INVESTIGATION OF HIV-1

ANTISENSE TRANSCRIPT EXPRESSION IN VIVO

by

Rachel Sklutuis

B.S. Biology (Towson University) 2016

THESIS

Submitted in partial satisfaction of the requirements

for the degree of

MASTER OF SCIENCE

in

BIOMEDICAL SCIENCE

in the

GRADUATE SCHOOL

of

HOOD COLLEGE

Spring 2022

Accepted:

_____	_____
Dr. Ann L Boyd, Ph.D. Committee Member	Ann L Boyd , Ph.D. Director, Biomedical Science Program
_____	
Dr. Fabio Romerio, Ph.D. Committee Member	
_____	_____
Dr. Mary Kearney, Ph.D. Thesis Adviser	April M. Boulton, Ph.D. Dean of the Graduate School

STATEMENT OF USE AND COPYRIGHT WAIVER

I do authorize Hood College to lend this thesis, or reproductions of it, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

DEDICATION

I would like to dedicate my thesis to my parents because without their love, support, and encouragement I would not be where I am today.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ann Boyd, Dr. Mary Kearney, and Dr. Fabio Romerio for being on my thesis committee and for supporting my project. I would also like to thank Dr. Jenn Groebner for all the daily time and guidance. I do not know how this project would have been completed without your help. Additionally, I need to thank Valerie Boltz for her assistance in planning this project and with generating a control transcript. Additionally, I would like to thank Val Turnquist, Connie Kinna, and Sue Toms for the administrative support. Finally, I would like to acknowledge all of the donors who made these studies possible as well as the funding sources including NCI intramural funding to MFK and funding from the Office of AIDS Research.

TABLE OF CONTENTS

ABSTRACT

Natural antisense transcripts (NATs) are a class of RNA molecules that function in regulating gene expression and are transcribed from the opposite strand of some protein-coding genes. In Human Immunodeficiency virus 1 (HIV-1), in vitro experiments have shown that expression of a NAT called *Ast* promotes the establishment and maintenance of viral latency through polycomb repressive complex 2 (PRC2)-induced epigenetic regulation of viral gene expression. The levels of *Ast* expression in single infected cells isolated from donors on suppressive antiretroviral therapy (ART) was investigated with a cell-associated RNA and DNA single-genome sequencing (CARD-SGS) assay adapted to detect *Ast* RNA. Additionally, a digital Polymerase Chain Reaction (PCR) approach was used to examine both sense and antisense HIV-1 gene expression. A median of ~5% of infected cells contained *Ast* RNA when measured by CARD-SGS. *Ast* expression levels ranged from 1-30 RNA copies/cell. *Ast* RNA measured using the digital reverse transcription PCR (RT-PCR) approach was detected in up to a median of 26% of infected peripheral blood mononuclear cells (PBMC). A similar fraction of infected cells contained sense *env* RNA as antisense *Ast* RNA in most donor samples. These findings confirm expression of HIV-1 *Ast* RNA in donors on ART. Detection of *Ast* RNA in unstimulated cells from donors on ART warrants the investigation of its bifunctional role as a coding and a regulatory RNA in vivo.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3TC	Lamivudine (nucleoside reverse transcriptase inhibitor)
ABC	Abacavir (nucleoside reverse transcriptase inhibitor)
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ASP	Antisense Protein
Ast	Antisense Transcript
ATL	Adult T cell Leukemia/Lymphoma
cDNA	Complementary DNA
CAR	Cell-associated RNA
CARD-SGS	Cell-associated RNA and DNA Single Genome Sequencing
DRV	Darunavir (protease inhibitor)
DTG	Dolutegravir (integrase inhibitor; also known as TCV)
Env	Envelope transcript
FTC	Emtricitabine (nucleoside reverse transcriptase inhibitor)
HAM	HTLV-1-Associated Myelopathy
HIV	Human Immunodeficiency Virus
HTLV	Human T cell Lymphotropic Virus
IN	Integrase
InR	Initiator Element
LRA	Latency Reversing Agent
NAT	Natural Antisense Transcript
NSP	Negative Sense Promoter

ORF	Open Reading Frame
PBMC	Peripheral Blood Mononuclear Cells
PCR	Polymerase Chain Reaction
PLWH	People Living With HIV
PRC2	Polycomb Repressive Complex 2
PSP	Positive Sense Promoter
RT-qPCR	Quantitative reverse transcription PCR
RTV	Ritonavir (protease Inhibitor)
TCV	Tivicay (integrase inhibitor; also known as DTG)
TDF	Tenofovir Disoproxil Fumarate (nucleotide analog)
TSP	Tropical Spastic Paraparesis

INTRODUCTION

HIV-1 Infection

Human Immunodeficiency Virus Type 1 (HIV-1) is a retrovirus that attacks the human immune system. When left untreated, HIV-1 leads to progressive loss of CD4+ T cells and immunological dysfunction that eventually results in the development of Acquired Immunodeficiency Syndrome (AIDS). AIDS is characterized by opportunistic infections and an increased likelihood of developing certain types of cancer (Fauci et al. 1984; Levine 1987; Longo et al. 1984). Since its discovery about 40 years ago, HIV-1 has been a major public health concern with 38 million people living with the virus today. Each year, 1.7 million people are newly infected with HIV-1 and 690,000 succumb to the infection (UNAIDS).

HIV-1 infection occurs when a mature virion attaches and fuses with a host target cell through interactions between the viral glycoprotein (Env) and the CD4+ receptor and CXCR4 or CCR5 coreceptors. Once fusion occurs, the viral capsid enters the cytoplasm of the cell and reverse transcription of the HIV-1 RNA genome begins as the capsid traffics to the host cell nucleus. In the nucleus, reverse transcription is completed, the capsid is uncoated and the viral enzyme integrase (IN) orchestrates integration of the viral DNA into the host genome (Burdick, et al, and Pathak 2019). The integrated proviral DNA is then able to be transcribed and translated by host cell machinery to produce viral proteins for assembly at the host cell membrane to generate new virus particles (Kleinpeter and Freed 2020) (**Figure 1)**. Some infected cells may have reduced or limited viral gene expression, allowing these cells to escape immune detection and persist as latently infected cells.

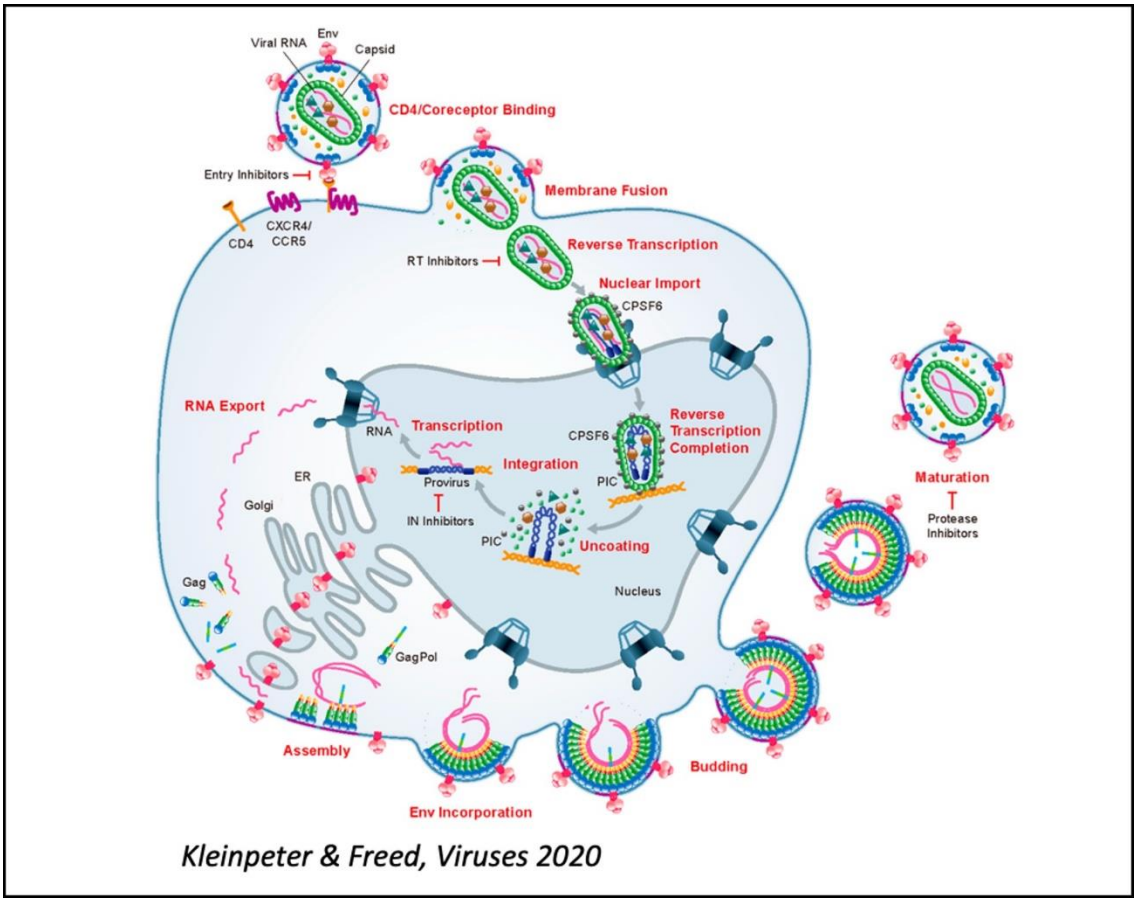


Figure 1: HIV-1 Replication Cycle adapted from Kleinpeter and Freed (Kleinpeter and Freed 2020) with permission.

HIV-1 Treatment

Currently, there is no widely available cure for HIV-1. However, antiretroviral therapy (ART) targets multiple steps in the virus replication cycle including entry, reverse transcription, integration, virus assembly, and maturation (Arts and Hazuda 2012). Although ART can effectively prevent new cells from becoming infected (Kearney et al. 2014; Lee et al. 2019; McManus et al. 2019; Van Zyl et al. 2017), low-level virus production from cells infected with intact HIV-1 genomes prior to the initiation of ART can persist (Palmer, et al. 2008). If therapy is interrupted, viral replication resumes, and viremia typically rebounds to pre-ART levels (Finzi et al. 1999; Finzi et al. 1997; Wong et

al. 1997). Despite the high efficacy of ART, only about 26 million people worldwide of the 38 million infected have access (UNAIDS). Although ART increases life expectancy of individuals with HIV-1, long-term treatment is associated with toxicities, metabolic complications, cardiovascular disease, kidney dysfunction, and bone loss (Chawla et al. 2018; HIV.gov 2019). Therefore, it is important to explore other mechanisms of treatment and potential cures for HIV-1.

One approach for a potential cure is a "kick and kill" strategy where the reservoir is stimulated to reverse viral latency so it can become targetable and killed by the individual's immune system. Although this strategy is attractive, progress on reversing viral latency in vivo has been limited (Pace and Frater 2019). A second possible approach is the "block and lock" strategy. This approach aims to permanently silence all proviruses by targeting the transcriptional machinery. Although this strategy is also appealing, it has yet to lead to an effective intervention for HIV-1 (Vansant et al. 2020).

HIV-1 Latency

Integration of HIV-1 into the host genome and persistence of latently infected cells present significant challenges to cure strategies. While the levels of virus decline about 100-fold to a viral set point after immune detection (Mellors et al. 1996) and to undetectable levels (<50 copies/ml) on ART (Raffi et al. 2017), the establishment of a latent reservoir occurs as early as 3 days into infection and is maintained by clonal expansion of infected cell populations (Maldarelli et al. 2014; Wagner et al. 2014). The latently infected cells that contain replication-competent proviruses comprise the reservoir that fuels rebound

viremia if ART is interrupted (Deeks et al. 2015). Therefore, understanding the mechanisms and maintenance of HIV-1 latency are critical for cure strategies.

There are many hypotheses about how HIV latency is established, but the exact mechanisms remain unknown (Dufour et al. 2020; Sengupta and Siliciano 2018). One thought is that integration into a non-genic region of the host DNA leads to less active transcription (Einkauf et al. 2019). Another is that histone modification through deacetylation and methylation can lead to transcriptional silencing (Archin et al. 2009; Friedman et al. 2011; Lusic et al. 2003). Additionally, methylation sites within the 5' LTR of the provirus have also been proposed to contribute to proviral silencing (Boltz et al. 2021; Kauder et al. 2009). A defect in RNA splicing has also been suggested to prevent virus production through incorrect splicing of viral proteins like Tat, which is required for expression of all viral transcripts (Dufour et al. 2020; Sengupta and Siliciano 2018). Lastly and of importance to studies proposed here, expression of an HIV antisense transcript (*Ast*) is thought to act as a long non-coding (lnc) RNA to promote and maintain proviral latency (Zapata et al. 2017).

Natural Antisense Transcripts (NATs)

Natural antisense transcripts (NATs) are RNA molecules transcribed from the opposite strand of a coding gene locus that can either be protein coding or non-coding (Khorkova et al. 2014). These transcripts have been identified in prokaryotes, eukaryotes, and viruses. Antisense transcripts can be divided into two categories, *cis* or *trans*, meaning that they act on sequences that originate from the opposite DNA strand of the same genomic locus or on transcripts originating from a different locus, respectively (Li et al. 2021; Rosikiewicz and Makalowska 2016) (**Figure 2**). NATs have been shown to have regulatory functions at the transcriptional level, at the post-transcriptional level, and through epigenetic changes.

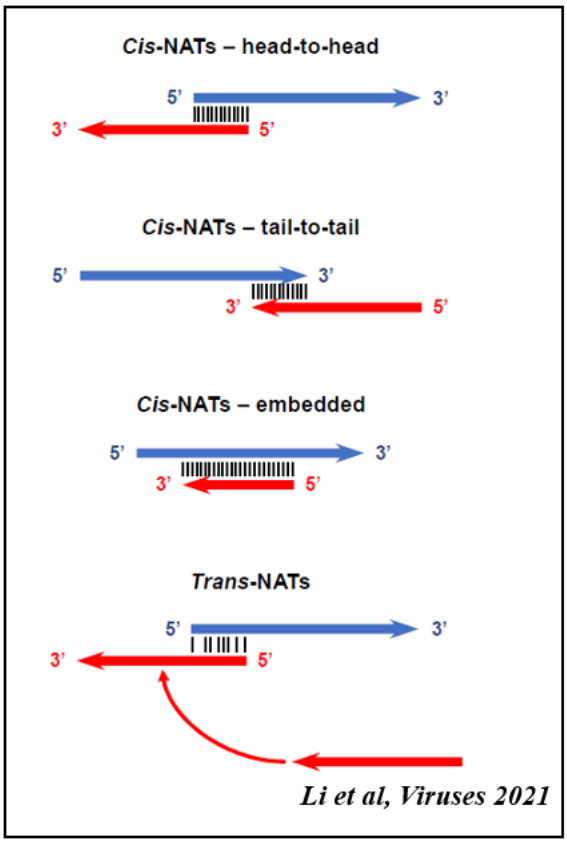


Figure 2: Types of Antisense Transcripts.

At the transcriptional level, NATs can inhibit sense gene expression through a several methods including promoter competition, binding site occlusion, RNA polymerase collision, and DNA and chromatin changes (Li et al. 2021; Zinad et al. 2017). Promoter competition occurs when sense and antisense RNA are expressed from a bidirectional promoter, and the assembly of the antisense transcriptional machinery blocks or prevents the formation of the sense transcriptional machinery. Binding site occlusion occurs when the passage of the RNA polymerase complex transcribing the NAT blocks access of the

transcription factors required for sense transcript expression to the chromatin. RNA polymerase collision occurs when transcriptional machinery is displaced by an RNA polymerase complex already assembled on the promoter or is stalled by another elongating transcriptional complex. (Li et al. 2021). Finally, DNA and chromatin changes that result in epigenetic silencing occur when the NATs are used as modular scaffolds to interact with DNA and proteins to generate specific functional complexes that promote epigenetic silencing of sense genes (Khorkova et al. 2014; Pelechano and Steinmetz 2013).

At the post-transcriptional level there are four ways that NATs can regulate expression of their paired sense transcript through the formation of double-stranded RNA. The first method is RNA masking which involves blocking the interaction of the sense transcript with proteins and miRNAs necessary for splicing stability, transport, and translation through the formation of a sense-antisense duplex (Wight and Werner 2013; Zinad et al. 2017). A second method called RNA interference involves the recognition of the RNA duplex by dicer, an enzyme that cleaves double stranded RNA (Zinad et al. 2017). A third mechanism involves the intracellular immune response triggered by the recognition of dsRNA molecules by protein kinase R and ultimately suppresses the protein expression. Finally, members of the ADAR (adenosine deaminase RNA-specific binding protein) protein family can recognize and deaminate the dsRNA molecules in a process called RNA editing (Li et al. 2021; Wight and Werner 2013).

NATs in Viral Systems

Natural antisense transcription has previously been documented in several eukaryotic viruses. Numerous studies have shown that antisense genes called latency-

associated transcripts (LATs), which are antisense to the immediate-early gene ICP0 in herpesvirus (HSV), are common (Stevens et al. 1987). The herpesvirus lifecycle is divided into latent and lytic stages. During lytic infection, virus is actively produced. Immediate early genes are expressed, which in turn regulate the expression of other genes. During the latent stage, LATs are abundantly expressed and have been shown to limit the accumulation of viral lytic gene transcripts during the establishment and maintenance of latency (Giordani et al. 2008). LATs are also thought to play a role in epigenetic regulation of HSV gene expression by heterochromatinization of lytic gene promoters (Wang et al. 2005). Finally, LATs may play a role in promoting latency reactivation by inhibiting apoptosis and promoting cell survival (Perng et al. 2000). A similar mechanism has also been identified in Varicella zoster virus (VZV) which has a latency-associated transcript (VLT) that lies antisense to the ICP0 homologue protein ORF61 and has been shown to suppresses the expression of ORF61 (Depledge et al. 2018).

Antisense transcripts have also been identified in a variety of other viruses including Kaposi's sarcoma-associated herpesvirus (KSHV) (Schifano et al. 2017), Epstein–Barr virus (EBV) (Majerciak et al. 2019), Marek's disease virus (MDV), bovine leukemia virus (BLV) (Durkin et al. 2016), simian T-leukemia virus type 1 (STLV-1) (Miura et al. 2013), murine leukemia virus (MLV) (Rasmussen et al. 2010), and bovine and feline immunodeficiency viruses (BIV and FIV) (Briquet et al. 2001; Liu et al. 2015).

Table 1: NATs in Viral Systems

Mechanism	NAT	Virus	Functions/Effects	References
Epigenetic Silencing	LATs	Herpesvirus (HSV)	Regulates lytic gene expression by promoting heterochromatinization during latency. May promote latency reactivation through inhibition of apoptosis and promotion of cell survival.	(Stevens et al. 1987) (Giordani et al. 2008) (Wang et al. 2005) (Perng et al. 2000)
	VLT	Varicella Zoster virus (VZV)	Regulates latency through the suppression of ORF61.	(Depledge et al. 2018)
Transcriptional Interference	HBZ	Human T cell leukemia virus 1 (HTLV-1)	Induces host genes involved in cell cycle progression and proliferation as well as anti-apoptosis factors, like survivin	(Clerc et al. 2008) (Matsuoka and Mesnard 2020)
Unknown	ALT	Kaposi's Sarcoma-associated herpesvirus (KSHV)	May play a role in regulating the viral lifecycle	(Schifano et al. 2017)
	EBNA-antisense transcript	Epstein-Barr virus (EBV)	Potentially regulates viral gene expression and plays a role in EBV reactivation from latency	(Majerciak et al. 2019)

NATs Encoded in HTLV-1

One of the best characterized NATs in a viral system is the *hbz* gene in human T-cell leukemia virus type 1 (HTLV-1). Similar to HIV-1, HTLV-1 is a retrovirus that predominantly infects CD4+ T cells. For 5-10% of individuals living with HTLV-1, the virus can lead to adult T cell leukemia/lymphoma (ATL) or HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). The interplay between *Hbz* (antisense) and *Tax* (sense) RNA expression and their protein products are thought to modulate the regulation of cellular pathways that promote survival and proliferation of

HTLV-1 infected cells (Bangham et al. 2019), influencing progression to either disease state. *Hbz* RNA expression is initiated at several positions within the R and U5 regions of the proviral 3'LTR, which lacks a TATA box (Cavanagh et al. 2006) and instead relies on Sp1, JunD, TCF1 and LEF1 promoter elements (Gazon et al. 2012). The *Hbz* transcript is also regulated at the epigenetic level through histone marks such as H3K9ac and H3K4me3 at the 3'LTR (Maeda et al. 1985). There are three major isoforms of *Hbz*, an unspliced and two alternatively spliced transcripts. During ATL, expression of one of the spliced transcripts is more prominent than the unspliced form and is thought to play a significant role in disease progression (Matsuoka and Green 2009), however both transcripts have been shown to exhibit similar functions in vitro (Ma et al. 2016).

Interplay between Tax and Hbz protein expression also influences disease progression. The TAX protein is a nuclear transcriptional transactivator that promotes sense strand transcription and functions to promote viral spread (Nyborg et al. 2010). Alternately, the HBZ protein inhibits sense strand transcription, suppressing TAX protein expression and preventing immune detection of infected cells (Bangham et al. 2019). The HBZ protein is stably expressed at low levels in a high proportion of cells while TAX is highly regulated and expressed intermittently (Billman et al. 2017; Satou et al. 2006). ATL has a clinically latent period that can last several decades and is associated with low levels of TAX protein expression and high expression levels of HBZ expression. The HBZ protein promotes latency by interacting with host factors that bind the viral cyclic AMP response elements in the 5'LTR like CREB, CREM, and ATF-1. This binding precludes the recruitment of TAX to the 5'LTR. Alternately, HAM/TSP is a chronic disease and is

associated with high HTLV-1 viral loads due to high TAX protein and low HBZ expression (Bangham et al. 2019; Philip et al. 2014).

A NAT Encoded in HIV-1

The existence of an antisense gene in HIV-1 (*asp*) was first reported in 1988 (Miller 1988). A highly conserved open reading frame (ORF) in the minus strand was found in the -2 reading frame which maps to same genomic region as the envelope gene at the junction between gp120 and gp41 (**Figure 3**). This ORF codes for a protein that is ~189 amino acids in length and is in a similar location to the HBZ reading frame in HTLV-1 (Manghera et al. 2017; Zapata et al. 2017). Antibodies recognizing the antisense protein (ASP) have been identified in cells collected from people living with HIV (PLWH) (Vanhee-Brossollet et al. 1995), and there is also evidence for a cytotoxic CD8+ T cell (CTL) response against ASP in vivo (Bet et al. 2015). More recently ASP has also been reported in HIV- expressing cell lines (Torresilla et al. 2015). ASP has been suggested to be incorporated into the membranes of budding virions as a structural element and/or to play a role in autophagy (Affram et al. 2019; Liu et al. 2019).

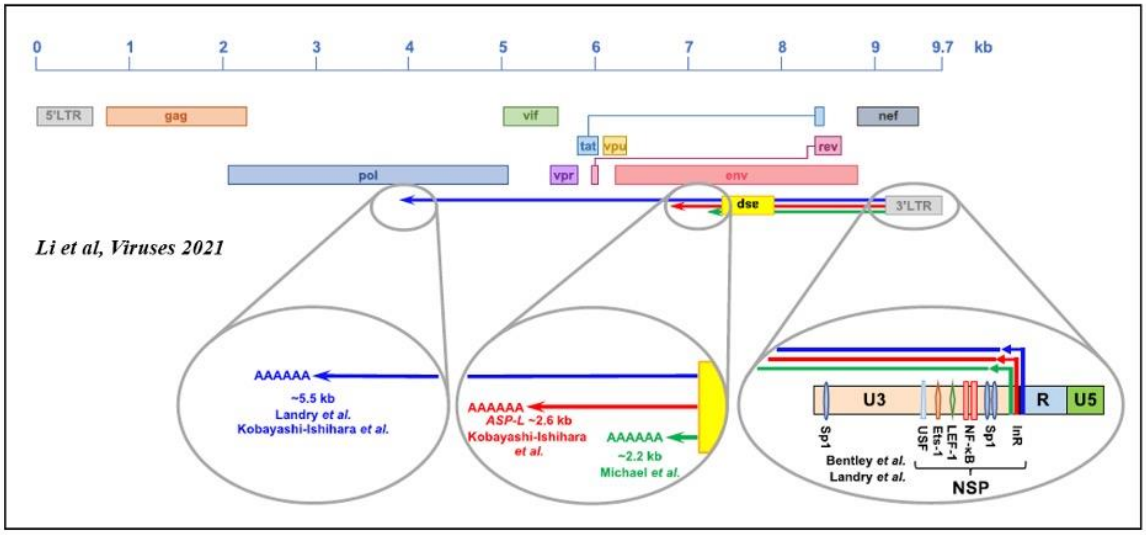


Figure 3: HIV-1 NATs Reported in the Literature.
Expression of HIV-1 NATs is dependent on a negative sense promoter within the U3 region of the 3'LTR. This is a TATA-less promoter that does not rely on Tat, but instead utilizes housekeeping genes and transcription factors like Sp1, NF-κB, LEF-1, Ets-1, and USF. The NAT start site is determined by initiator elements (InR) close to the U3-R boundary. Several different research groups have described the length as varying between 2.2 and 5.5 kb.

Northern blot analysis of Poly-A+ RNA extracted from H9 cells acutely infected with HIV-1 strain IIIB provided some of the first experimental evidence for an antisense gene. This study demonstrated that that antisense transcription was restricted to the early phases of acute infection (Bukrinsky and Etkin 1990). The start site, length and polyadenylation site of HIV-1 antisense transcription have been studied by many groups in several HIV-1 infected cell lines. These studies have shown that the HIV-1 proviral genome does transcribe an antisense transcript that can produce one or more species of NATs of various lengths that can be spliced or unspliced. They have also provided evidence that the 5' terminus of the antisense transcript maps within the 3'LTR, demonstrating that these antisense RNA molecules are encoded within the provirus and not the product of read-through transcription initiated from downstream cellular promoters. Although there are some inconsistencies in the reports from various groups, these differences can be

explained by several factors including the cell systems, the replication competence of the virus, stage of infection, and techniques used.

A number of studies have found that antisense transcription in HIV-1 is driven by a negative sense promoter (NSP) located in the U3 region of the 3'LTR and that this promoter relies on both ubiquitous (Sp1, LEF-1, USF and Ets-1) and inducible (NF-κB) transcription factors. This promoter has been found to lack a TATA box, and instead utilizes an initiator element (InR) to determine the transcription start site (**Figure 3**). The NSP has also been shown to be weaker than the positive sense promoter (PSP), and potentially inhibited by Tat expression because the transcriptional machinery is being directed to the PSP (Bentley et al. 2004; Landry et al. 2007).

Cassan et al. evaluated a dataset of HIV-1 and SIV-1 sequences to study the origin, evolution, and conservation of the *asp* gene. They found that the *asp* open reading frame is specific for HIV-1 group M, the strain responsible for the pandemic in humans. Through computer simulations and the analysis of stop codons, they demonstrated that the conservation of the ORF could not be due to chance and that *asp* does impose a selection pressure (Cassan et al. 2016). Further statistical analysis by Pavesi and Romerio (2022) also concluded that the emergence of the *asp* gene was a recent de novo development that arose from the overprinting of the env gene. This process involved the creation of a start codon, the gradual removal of existing internal stop codons, as well as nucleotide sequence evolution that made the emergence of new stop codons unlikely. Through this analysis as well as the observation that this open reading frame has been conserved in the pandemic group M strain of HIV-1, they concluded that the *asp* provides a selection advantage to HIV-1 (Pavesi and Romerio 2022).

HIV-1 *Ast* has been shown to inhibit viral replication and promote and maintain latency in stably expressing cell lines (Kobayashi-Ishihara et al. 2018; Zapata et al. 2017). The antisense transcript has been shown to be inefficiently polyadenylated and to be predominately retained in the nucleus to act as a lnc RNA (Ma et al. 2021). More specifically, in vitro studies showed that *Ast* recruits and retains the enhancer of zeste homolog 2 (EZH2), a core component of the polycomb repressive complex 2 (PRC2), to the HIV-1 5'LTR. The recruitment of the EZH2 catalyzes trimethylation of lysine 27 on histone H3 (H3K27me3), a supressive epigenetic mark that promotes nucleosome assembly and suppression of viral transcription (**Figure 4**). This mechanism suggests that *Ast* acts as a lnc RNA to promote epigenetic silencing of the HIV-1 5'LTR, in order to induce and maintain viral latency in HIV-1 infected cells (Zapata et al. 2017).

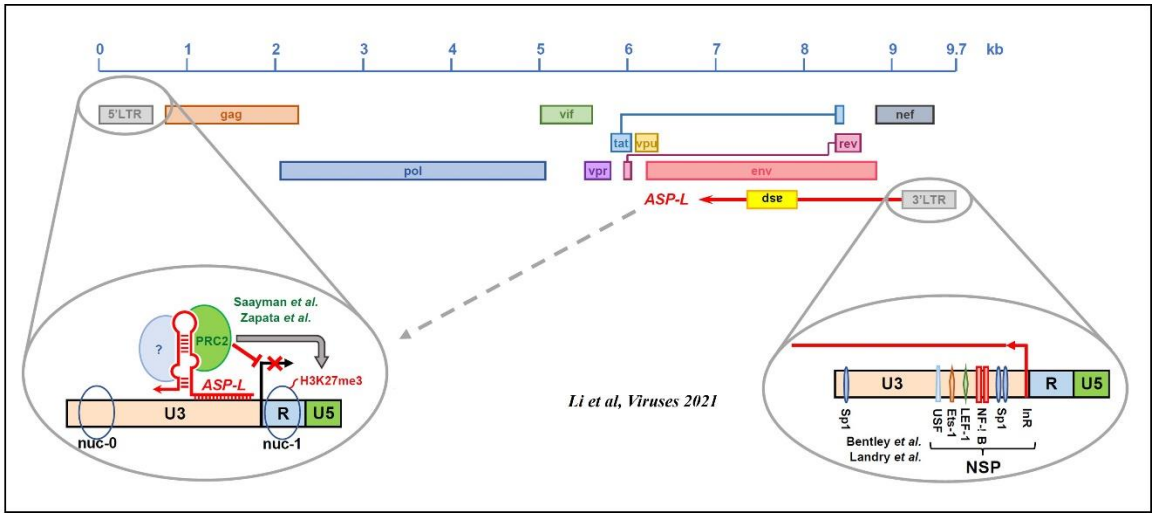


Figure 4: *Ast* Role in HIV-1 Latency.
HIV-1 *Ast* contributes to the establishment and maintenance of HIV-1 latency by negatively regulating transcription at the 5'LTR. *Ast* recruits chromatin-modifying enzymes, DNA methyltransferase 3a and histone methyltransferase EZH2, core components of the polycomb repressor complex 2 (PCR2), to the 5'LTR which catalyze methylation of histone 3.

While the existence of *Ast* and its role as a potential mechanism for latency have been examined in vitro in a variety of cell lines, few studies have examined *Ast* in HIV-1 infected cells from donors on ART. Briefly, Zapata et al. (2017) examined the expression levels of *Ast* in CD4+ T cells from three donors with HIV-1 on ART (for greater than 24 months) and two uninfected donors using a strand-specific quantitative RT-PCR (RT-qPCR) assay. *Ast* was detected only in infected cells from the three donors with HIV-1. *Ast* levels ranged from 10 to 30 copies per million CD4+ T cells (Zapata et al. 2017). Another study by Mancarella et al. (2019) examined *Ast* expression in CD4+ T cells from untreated viremic donors with HIV-1 as well as from donors on ART. *Ast* was detected in both groups of donors, and similar levels of *Ast* expression in infected cells from donors on ART were seen as in those reported by Zapata et al. However, in order to detect *Ast*, CD4+ T cells were stimulated ex vivo with IL-2 and plated in anti-CD3/CD28-coated 48 well plates. Stimulations were carried out for 4-5 days and samples were assayed daily. *Ast* RNA copies ranged from 30 to 2 million per million CD4+ T cells during peak timepoints during 5 day stimulations (Mancarella et al. 2019).

The studies from these two groups (Mancarella et al. 2019; Zapata et al. 2017) are important contributions but have several limitations. Mancarella et al. used ex vivo stimulation of donor CD4+ T cells, which may not accurately reflect *Ast* expression in vivo. Additionally, they used a biotinylated specific antisense primer for reverse transcription followed by purification of the biotinylated cDNA, which may affect the overall efficiency of their assay. Neither of these studies sequenced the cDNA generated in the assay to ensure that their amplification products were HIV-1 *Ast* and both assessed the cDNA in bulk rather than at the single-cell level. Therefore, there is a need to investigate *Ast* expression in single

HIV-1 infected cells from donors on ART. If so, what fraction of HIV-1 infected cells express *Ast* and to what levels? How do the levels of *Ast* expression compare to the expression of other HIV-1 genes? What is the length of the antisense transcript? Lastly, are the levels and fraction of HIV-1 infected cells with *Ast* different in donors on ART vs during antiretroviral therapy interruption (ATI)? Understanding the expression of *Ast* in vivo will contribute to our understanding of HIV-1 persistence in donors on ART.

To answer these questions, the cell-associated RNA and DNA single-genome sequencing assay (CARD-SGS) was adapted from Wiegand et al. to detect *Ast* as well as antisense transcripts of other genes throughout the provirus. Additionally, a digital PCR approach was developed to examine both sense and antisense expression in the same region of envelope. Both assays were then used to investigate the expression of these antisense transcripts in PBMC and CD4+ T cells from donors on long-term suppressive ART.

MATERIALS AND METHODS

Donor Samples

Samples from 4 HIV-1 infected participants were obtained from the San Francisco SCOPE cohort (NCT00187512) (**Table 2**). PBMCs were separated with Ficoll and resuspended in FBS with 10% DMSO. In addition, peripheral blood cells obtained by leukapheresis from PID 3162 were also enriched for CD4+ T cells before freezing at -80C/LN. All samples were shipped on dry ice and stored in liquid nitrogen until processed.

Table 2: Donor Characteristics

PID	Age	Sex	Race/ Ethnicity	Duration of Infection (yrs)	Duration on ART (yrs)	Duration of continuous viral suppression on ART at sampling	Current ART Regimen
2669	56	M	White Non-Hispanic	Unknown	<15.8	4.3 years (T1); 5.5 years (T2); 2 weeks* (T3); 1 month* (T4)	ABC/3TC /DTG
1079	64	M	Latino	17.0	13.0	12.8 years	FTC/TDF , ETV
1683	45	M	White Non-Hispanic	7.7	5.7	5.4 years	FTC/TDF , RTV, DRV
3162	57	M	White Non-Hispanic	30.0	19.0	11.2 years	RTV, DRV, ABC/3TC /TCV

* Due to ART interruption/nonadherence, PID 2669 was not suppressed at timepoint (T3), but was suppressed at T4.

Sorting Donor T Cells

For PID 3162, PBMC or enriched CD4+ T cells were sorted using fluorescence-activated cell sorting (FACS) to obtain HLA-DR- (resting) CD4+ T cells by the Vaccine Research Center (VRC) at the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD (**Figure 5)**.

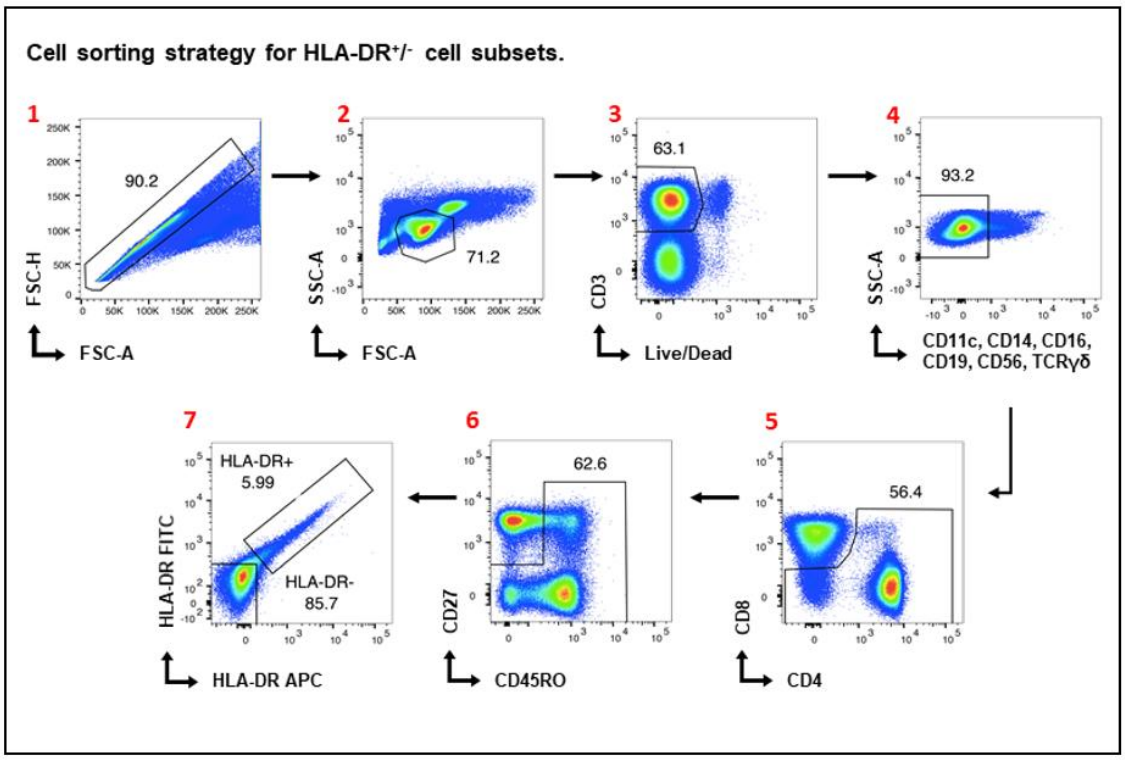


Figure 5: Fluorescence Activated Cell Sorting (FACS) of CD4+ T cells.
PBMC and/or enriched CD4+ T cells were sorted into resting (HLA-DR-) or activated (HLA-DR+) T cell subsets as follows: **1)** Singlet gate to count single cells and exclude cell doublets or aggregates. **2)** Lymphocyte gate is to exclude monocytes and granulocytes. **3)** CD3 T cells and live/dead gate is to exclude non-T cells and any dead cell. **4)** Dump gate is to exclude cells from analysis that stain positive for lineage markers not associated with CD4+ T cells. In this case dendritic cells (CD11c+), monocytes (CD14+), NK cells (CD16+CD56+), B cells (CD19+) and T cells with a gamma delta receptor (TCR$\gamma\delta$+) are excluded. **5**) CD4+ gate to exclude the majority of CD8 T cells. **6**) Memory gate to exclude naïve CD4+ T cells from memory CD4+ T cells. **7**) Memory CD4+ T cells were gated for HLA-DR expression to sort for HLA-DR+ or activated memory CD4+ T cell (upper gate) and HLA-DR- or resting CD4+ T cells (lower gate).

Briefly, cells were stained with LIVE/DEAD Aqua statin (ThermoFisher Scientific, L34965) and the following antibodies (see Table 3): CD3_APC-H7, CD11c-Cy5.5PerCP, CD14-Cy5.5PerCP, CD16-Cy5.5PerCP, CD56-Cy5.5PerCP, CD19-Cy5.5PerCP, TCR γδ-Cy5.5PerCP, CD4+-BV785, CD8α-BV711, CD27-Cy5PE, CD4+5RO-ECD, HLA-DR APC and HLA-DR FITC. Cells were first sorted by forward scatter and gated for cell singlets to ensure that doublets or aggregates are excluded. Single cells were then sorted by forward and side scatter and lymphocytes were gated to exclude monocytes and granulocytes. Lymphocytes were gated for live CD3+ T cells next to ensure non-T cells and any dead cells are excluded using a LIVE/DEAD Aqua statin (ThermoFisher Scientific, L34965). A dump gate was used to exclude dendritic cells (CD11c+), monocytes (CD14+), NK cells (CD16+/CD56+), B cells (CD19+) and T cells with a gamma delta receptor (TCR γδ+) from the gated population. CD3+ T cells were then sorted and gated on CD4+CD8- cells to exclude the majority of CD8+ T cells. This gate selected for any CD4+ T cells including CD4+ T cells also positive for CD8+ since there is a potential for them to be infected. CD4+- or CD8- T cells are also collected since CD4+ is downregulated with HIV infection, so some of those cells could also be positive for HIV. Next, CD4+ T cells are gated for memory T cells (CD27-CD4+5RO- and CD27-CD4+5RO+) and further sorted into HLA-DR+ (activated) or HLA-DR- (resting) CD4+ memory T cells. HLA-DR expression is known to be induced in CD4+ T cells upon cell activation. Naïve CD4+ T cells (CD27+CD4+5RO-) T cells were also collected (**Figure 5, step 6)**.

Table 3: Cell Sorting Antibodies

Antibody	Clone	Catalog Number
CD3-APC-H7	SK7	BD Biosciences, Cat# 641406
CD4+-BV785	OKT4	BioLegend, Cat# 317442
CD8α-BV711	RPA-T8	BioLegend, Cat# 301044
CD11c-Cy5.5PerCP	S-HCL-3	BioLegend, Cat# 371518
CD14- Cy5.5PerCP	M5E2	BioLegend, Cat# 301824
CD16-Cy5.5PerCP	3G8	BioLegend, Cat# 302028
CD19-Cy5.5PerCP	SJ25C1	BioLegend, Cat# 363016
CD27-PE/Cy5	1A4CD27	Beckman Coulter, Cat# 6607107
CD4+5RO-ECD	UCHL1	Beckman Coulter, Cat# IM2712U
CD56-Cy5.5PerCP	HCD56	BioLegend, Cat# 318322
CD57-BV421	NK-1	BD Biosciences, Cat# 563896
CCR7-A700	150503	BD Biosciences, Cat# 561143
TCR γδ-Cy5.5PerCP	B1	BioLegend, Cat# 331224
HLA-DR-FITC	L243	BioLegend, Cat# 307604
HLA-DR-APC	G46-6	BD Biosciences, Cat# 559866

Cell Lines Used for Assay Development

ACH2 cells, a human T cell model for HIV latency, were used for assay development. They are chronically infected and have a single integrated copy of the HIV-1 strain LAI and produce only low levels of virus. CEM cells, a human T lymphoblast cell line, were also used as an uninfected control to provide a background for infected cells.

Both cell lines were grown in suspension at 37ºC and passaged approximately every three days to give a concentration of $1x10^6$ cells/mL. The cells were grown in RPMI 1640 (Gibco, 21870076) supplemented with 1% Penicillin-Streptomycin-Glutamine (Gibco, 10378016) and 10% FBS (Gibco, 10438018). Cells were cryopreserved in aliquots of $1x10^6$ cells and 10% DMSO (Sigma, D2650) in a volume of 1mL medium and stored in liquid nitrogen until needed. ACH2 cells were spiked into uninfected CEM cells for assay development to reflect frequencies of infected:uninfected T cells in human samples. RPMI medium (Lonza #12-702F) was warmed to 37°C and added drop-wise to thawed cells.

Cells were then divided into aliquots and centrifuged at 500×g for 5 minutes to pellet. The supernatant was removed with a small transfer pipette and the pelleted cells were used immediately or were frozen on dry ice and stored in liquid nitrogen until ready for use.

Donor Sample Preparation

RPMI medium (Lonza #12-702F) was warmed to 37°C and added dropwise to thaw PBMCs for nucleic acid extraction. Thawed donor PBMCs were divided into aliquots of $5x10^5$ or $1x10^6$ cells and centrifuged at 500×g for 5 min. The supernatant was removed with a small transfer pipette and the pelleted cells were used immediately for downstream assays or were frozen on dry ice and stored in liquid nitrogen until ready for use.

Genomic DNA and RNA Extraction

Total nucleic acid was extracted following the method described in Wiegand et al., (Wiegand et al. 2017) by adding 100µL of 3M guanidine HCl and 5µL of 20mg/mL Proteinase K (Applied Biosystems #AM2548) to each cell pellet. The samples were vortexed for 10 seconds to resuspend the pellet and incubated at 42°C for 1 hour. Next, 400µL of 6M guanidine isothiocyanate (GuSCN) and 8µL of 20mg/mL glycogen (Roche #10901393001) was added to each sample, mixed, and incubated at 42°C for an additional 10 minutes. Nucleic acids were precipitated by adding 500µL of 100% isopropanol and vertexing well, followed by centrifugation at 21,000×g at room temperature for 10 minutes. The supernatant was removed, and the pellet was washed with 750µL of 70% ethanol, followed by additional centrifugation at 21,000×g for 10 minutes. Precipitated nucleic acid pellets were air-dried for 5 minutes and 100µL of 5mM Tris·HCl (pH 8.0) was added to

resuspend DNA. DNA was incubated at 42°C for 2 hours to further facilitate resuspension and then stored at -80°C until needed.

For RNA isolation, the extracted nucleic acid pellet was resuspended in 38µL of DNase buffer and 2µL of 10 units/µL DNase I (Roche #04716728001) and incubated for 20 minutes at 37°C. After incubation, 200 µL of 6M GuSCN (Sigma #50983) was added and mixed well, followed by the addition of 250µL of 100% isopropanol. Samples were vortexed for 10 seconds, then centrifuged at 21,000×g for 10 minutes to pellet nucleic acid. Supernatant was removed and pellets were washed with 750µL of 70% ethanol, followed by additional centrifugation at 21,000×g for 10 minutes. Finally, the RNA pellet was air-dried, then resuspended in 20 µL of 5 mM Tris·HCl (pH 8.0) and used immediately for cDNA synthesis or stored at -80°C.

cDNA Synthesis

The cDNA was synthesized by adding 2.5µL of 10mM dNTPs and 2.5µL of 2µM gene-specific tagged cDNA primer (**Table 4**) to each 20µL RNA sample. The RNA was denatured at 85°C for 10 minutes, then immediately cooled at -20°C for 1 minute. Next, 25µL of SuperScript III Reverse Transcriptase (Invitrogen #18080-044) master mix was added to the denatured RNA. Master mix was made by combining 10µL of 5X first strand buffer, 0.5µL of 0.1M DTT, 13.5µL of RNase-free water, 0.5µL of 40 U/µL RNaseOUT recombinant ribonuclease inhibitor (Invitrogen #10777-019), and 0.5µL of 200 U/µL SuperScript III Reverse Transcriptase. cDNA was synthesized at 45°C for 1 hour, then cooled to 4°C. RNA was removed from the cDNA sample by adding 1µL RNase H (5units) (NEB #M0297S) and incubating at 37°C for 20 minutes. Finally, enzymes were inactivated

at 75°C for 10 minutes. The cDNA is then cooled to 4°C and used immediately for downstream assays, stored at -80°C, or ethanol precipitated before being used for further analyses.

Table 4: cDNA Synthesis Primers

Primer Name	Primer Sequence (5'-…-3')	Primer Use
AST-Tagged Primer	CTGATCTAGAGGTACCGGATCCAACATGTGGCAGAAAGTAGG	*Ast*
ENV-Tagged Primer	CTGATCTAGAGGTACCGGATCCACATCTAATTTGTCCACTGA	*Env*
1849Tag+	CTGATCTAGAGGTACCGGATCCGATGACAGCATGTCAGGGAG	Antisense P6-RT
3996Tag+	CTGATCTAGAGGTACCGGATCCCATCTAGCTTTGCAGGATTCG	Antisense IN
G00-Tag	CTGATCTAGAGGTACCGGATCCGACTAGCGGAGGCTAGAAG	Antisense Gag

Ethanol Precipitation

cDNA was transferred to a low bind Eppendorf tube and 0.1 vol 3M Sodium Acetate pH 5.5 (Invitrogen #AM9740) and 1μL 20mg/ml glycogen (Roche # 1090139300) was added. After mixing, 3 vol 95% ETOH was added, and the entire mixture was vortexed for 10 seconds and incubated overnight at -20°C. The next day, the precipitated cDNA was centrifuged at 21,000xg for 20 minutes at room temperature, and all supernatant was removed. The pellet was washed in 600μL 70% ETOH, pulse vortexed, and centrifuged for 15 minutes at 21,000xg at room temperature. All supernatant was removed and the cDNA pellet was air dried and then resuspended in 20μL of 5mM Tris·HCl (pH 8.0). The cDNA was then used immediately or stored at -80°C

Quantification of the Number of HIV-1 Infected Cells

Extracted DNA was assessed for the number of integrated proviruses per million cells using a real-time PCR method performed using the Lightcycler 480 Probes Master hot start reaction mix (Roche #4707494001) in 20μL volumes per reaction. A 2mL master mix was made by combining 1mL of 2X Lightcycler 480 Probes Master hot start reaction mix, 12μL of 100μM U5 Fwd primer (5'- CTTAAGCCTCAATAAAGCTTGCC-3'), 12μL of 100μM U5 Rev primer (5'- GGATCTCTAGTTACCAGAGTC-3'), 2μL of 100μM U5 probe (5'-/5HEX/TTCCTTGGGTTCTTGGGAGCAG/ZEN//3IaBkFQ/-3'), 774μL of RNase-free water, and 200μL of diluted DNA. Reactions were carried out at an estimated endpoint of <30 copies of DNA spread across a 96-well plate. Estimates were calculated with previously known information about these donors from McManus et.al. (McManus et al. 2019). The calculated amount of DNA was diluted in 5mM Tris·HCl (pH 8.0) to bring the final volume of sample up to 200μL. The master mix was then spread across a 96-well plate with 20μL per well. The quantitative PCR reactions were carried out on a Roche LightCycler480 at a denaturation temperature of 95ºC for 10 min and 55 cycles at 60ºC. Two negative (no template) and two positive (ACH2 cell DNA) controls were included on each PCR plate. The number of positive wells was used to calculate the number of infected cells per million in each of the donor samples.

Cell-Associated RNA and DNA-Single Genome Sequencing (CARD-SGS)

CARD-SGS was modified from the protocol described in Wiegand et.al. (Wiegand et al. 2017) and was used to sequence and quantify HIV DNA and sense and antisense HIV RNA in *env*, *gag*, and *pol*. Briefly, cDNA or DNA was diluted to a near endpoint in a

volume of 200µL and added to 800µL of Platinum II Taq (Thermo Fisher #14966025) master mix containing the following: 200µL of 5X Platinum II PCR Buffer, 20µL of 10mM dNTPs, 4µL of each 50µM primer (**Table 5**), 564µL of molecular-grade water, and 8µL of Platinum II Taq polymerase. The total volume of cDNA and master mix was spread across a 96-well plate (10µL per well). PCR cycling was performed as described in **Table 6**. Following the first round of PCR, each well was diluted with 50µL of 5mM Tris·HCl (pH 8.0). For the nested PCR, 2µL of PCR 1 product was transferred from each well to a 96-well plate containing 8µL of Platinum II Taq master mix (described above). Nested PCR was performed with the cycling conditions in **Table 6**. Positive wells were identified using GelRed detection to illuminate amplified DNA (Biotium #41003). Briefly, 4µL of the GelRed is diluted in 10mL of Tris·HCl (pH 8.0). Next, 15µL of the diluted GelRed is then added to each well of a 96-well plate along with 15µL of the diluted PCR product. The plate is vortexed, centrifuged, then read using UV gel imager. Positive PCR reactions were sequenced by Sanger.

Table 5: CARD-SGS Primers

Primer Name	Primer Sequence (5'-…-3')	Primer Use
Tag Fwd	CTGATCTAGAGGTACCGGAT	PCR 1 Fwd (When tagged cDNA primer used)
AST-OR	TGGTACTAGCTTGTAGCACCATCC	*Ast* PCR1
AST-IF	AGCAGAACAATTTGCTGAGGGC	*Ast* PCR 2
AST-IR	GTCATTGGTCTTAAAGGTACCTGAGG	
1849+	GGATCCGATGACAGCATGTCAGGGAG	P6-RT PCR1
3500-	CTATTAAGTCTTTTGATGGGTCATAA	
1870+	GAGTTTTGGCTGAAGCAATGAG	P6-RT PCR2
3410-	CTGTTAGTGGTATTACTTCTGTTAGTGCTT	
5270-	CTGACCCAAATGCCAGTCTC	IN PCR1
4133+	GGAAAAGGTCTATCTGGCATG	IN PCR2
5248-	TCTCCTGTATGCAGACCCCA	
G01	AGGGGTCGTTGCCAAAGA	Gag PCR1
G10	CAGTATTAAGCGGGGGAGAATT	Gag PCR1
G15	CTTTGCCACAATTGAAACACTT	
envB5out	TAGAGCCCTGGAAGCATCCAGGAAGT	Env (to design donor-specific primers)
9538r	AGAGAGACCCAGTACAGGCAAAA	
envB5in	TTAGGCATCTCCTATGGCAGGAAGAAG	
9418r	CAAGCTCGATGTCAGCAGTTCT	

Table 6: CARD-SGS PCR Cycling Conditions

Assay	PCR 1	PCR2
Ast CARD-SGS	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 30 s 5. Go to #2, 44 cycles 6. 68ºC for 1 min 7. 4ºC hold	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 20 s 5. Go to #2, 39 cycles 6. 68ºC for 1 min 7. 4ºC hold
P6-RT CARD-SGS (Sense and Antisense)	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 30 s 5. Go to #2, 44 cycles 6. 68ºC for 1 min 7. 4ºC hold	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 30 s 5. Go to #2, 39 cycles 6. 68ºC for 1 min 7. 4ºC hold
Antisense Gag CARD-SGS	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 25 s 5. Go to #2, 44 cycles 6. 68ºC for 1 min 7. 4ºC hold	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 20 s 5. Go to #2, 39 cycles 6. 68ºC for 1 min 7. 4ºC hold
Antisense IN CARD-SGS	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 20 s 5. Go to #2, 44 cycles 6. 68ºC for 1 min 7. 4ºC hold	1. 94ºC for 2 min 2. 94ºC for 15 s 3. 60ºC for 15 s 4. 68ºC for 20 s 5. Go to #2, 39 cycles 6. 68ºC for 1 min 7. 4ºC hold

Sequence Analysis

Sequences obtained by Sanger were aligned using ClustalW which generates alignments based off pairwise genetic distances. Population genetic diversity was calculated as average pairwise distance (APD) using MEGA 7.0 (Kumar et al. 2016). Neighbor joining trees were also constructed using MEGA 7.0 and rooted on a consensus HIV-1 Subtype B sequence. Identical sequences obtained from the same cell extraction aliquot were considered to have resulted from a single infected cell. Detection and analysis

of APOBEC-induced hypermutations were determined by using the online Los Alamos National Laboratory HIV sequence database tool (https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html).

Table 7: Sequencing Primers

Primer Name	Primer Sequence (5'-…-3')	Primer Use
AST-Seq2 Rev	GGTGAATATCCCTGCCTAACTCTAT-3'	*Ast*
AST-Seq2 Fwd	GGTTTAACATAACAAATTGGCTGTGGTATATAA	
AST-Seq3 Fwd	ATGGGTGGCAAGTGGTCAAA	
AST-IF	AGCAGAACAATTTGCTGAGGGC	
AST-IR	GTCATTGGTCTTAAAGGTACCTGAGG	
2030+	TGTTGGAAATGTGGAAAGGAAGGAC	P6-RT
2600+	ATGGCCCAAAAGTTAAACAATGGC	
2610-	TTCTTCTGTCAATGGCCATTGTTTAAC	
3330-	TTGCCCAATTCAATTTTCCCACTAA	
Poli9D	AAAATTAGCAGGAMGATGGCCAG	IN
Poli10B	TATTCATAGATTCYACTACTCCTTG	
4133+	GGAAAAGGTCTATCTGGCATG	
5248-	TCTCCTGTATGCAGACCCCA	
G30	CAGTAGCAACCCTCTATTGTGT	Gag
G25	ATTGCTTCAGCCAAAACTCTTGC	
G10	CAGTATTAAGCGGGGGAGAATT	
G15	CTTTGCCACAATTGAAACACTT	
For16	TTTAATTGTGGAGGAGAATTTTTCTA	Env
For17	AGCAGCAGGAAGCAGTATGGGCGC	
For18	CATATCAAATTGGCTGTGGTATAT	
Rev15	CTGCCATTTAACAGCAGTTGAGTTGA	
Rev16	ATGGGAGGGGCATACATTGCT	
envB5in	TTAGGCATCTCCTATGGCAGGAAGAAG	
envB5out	TAGAGCCCTGGAAGCATCCAGGAAGT	
9418r	CAAGCTCGATGTCAGCAGTTCT	

Digital PCR (Endpoint PCR with Primers and Probes)

Synthesized and ethanol precipitated cDNA was also assessed for *Ast* and *Env* expression by a RT-PCR method using the Lightcycler 480 Probes Master hot start reaction mix (Roche Diagnostics) in 20µL volumes per reaction. A 2mL master mix was made as previously described to include 200µL of diluted cDNA (**Table 8)**. Reactions were carried out at an estimated endpoint of <30 copies of cDNA spread across a 96-well plate. Estimates were calculated by considering the number of infected cells per million and the results from the CARD-SGS assay. A calculated amount of cDNA was used with 5mM Tris·HCl (pH 8.0) to bring the final volume of sample up to 200µL. The master mix was then spread across a 96-well plate with 20µL per well. The reaction was carried out on a Roche LightCycler 480 at a denaturation temperature of 95ºC for 10 min and 55 cycles at 60ºC. The number of positive wells was divided by the number of infected cells assayed to calculate the percent of infected cells expressing either *env* or *Ast* RNA.

Table 8: Digital PCR Primers and Probes

Primer Name	Primer/Probe Sequence (5'-…-3')	Primer Use
AST-Tag (shifted)	ATCTAGAGGTACCGGATCCAAC	*Ast*
AST Rev-qPCR	TGATGAACATCTAATTTGTCCACTGA	
AST Probe	/56-FAM/AGCAATGTATGCCCCTCCCA/ZEN//3IaBkFQ/	
ENV-Tag (shifted)	ATCTAGAGGTACCGGATCCACA	*Env*
ENV Fwd-qPCR	ACAAATTATAAACATGTGGCAGAAAGTAGG	
Env Probe	/56-FAM/TGGGAGGGGCATACATTGCT/ZEN//3IaBkFQ/	

Generation of Control RNA and DNA

Since both CARD-SGS and digital PCR were to be performed on cDNA synthesized with a tagged-primer, a cDNA control with the tag on the 5' end needed to be generated to determine the efficiency of the PCR reactions. To accomplish this, ACH2 DNA was amplified using the AST-Tagged primer (**Table 4**) and AST-OF primer (**Table 5**) using the *AST* CARD-SGS PCR1 cycling conditions described in **Table 6**. This control DNA with the tag sequence at the 5' end was then diluted to ~1 copy/μL and used as a control template in CARD-SGS and digital PCR reactions.

AST was amplified from a pUC57-AST plasmid obtained from Fabio Romero (Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD) using M13 fwd (5'-GTAAAACGACGGCCAGT-3') and M13 rev primers (5'-CAGGAAACAGCTATGAC-3'). Using Platinum II Taq (Thermo Fisher: Cat#14966025), 50μL PCR reactions were carried out with 50ng of plasmid DNA and 10μL of 5X Platinum II PCR Buffer, 1μL of 10mM dNTPs, 1μL of each 10μM primer, 0.4μL of Platinum II Taq polymerase, and molecular-grade water to bring to 50μL. PCR cycling was performed as follows: 94°C for 2 min and 45 cycles of 94°C for 15 s, 60°C for 15 s, and 68°C for 45 s, followed by a final extension at 68°C for 1 minute. The PCR product was then purified using a QIAquick PCR Purification Kit (Qiagen # 28104). Using a PCR cloning kit (New England Biolabs #E1202s), 80ng of the purified PCR product was ligated into a linearized pMiniT 2.0 vector and transformed into stable competent E.coli cells following kit instructions (**Figure 6)**. Transformed colonies were picked, grown overnight, and miniprepped using a QuickLyse Miniprep Kit (Qiagen #27406). Plasmids were screened for the insert using the provided cloning analysis forward

(5'-ACCTGCCAACCAAAGCGAGAAC-3') and reverse (5'-TCAGGGTTATTGTCTCATGAGCG-3') primers in the PCR cloning kit and cycling conditions as described above. Plasmid were then sequenced by Sanger (**Table 9**) to confirm orientation and sequence.

Figure 6: In vitro Expression of *Ast* Using pMiniT Expression Vector.

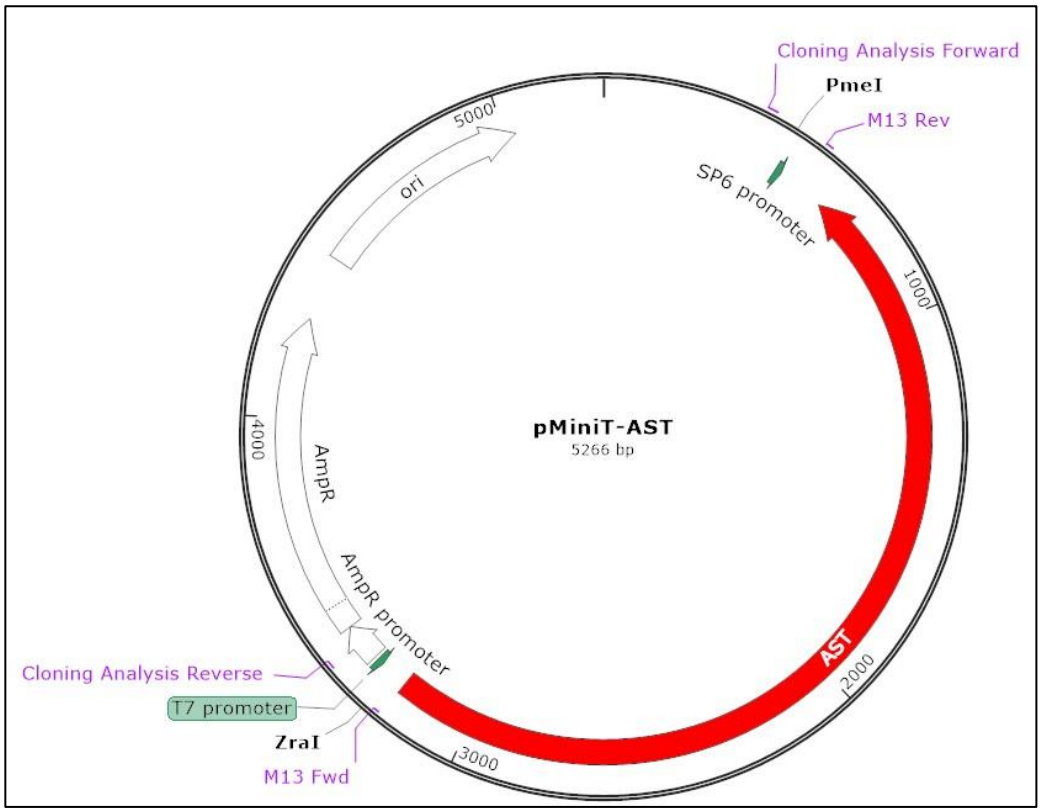


Table 9: Plasmid Sequencing Primers

Primer Name	Primer Sequence (5'-…-3')
AST-NEF2R	GTCATTGGTCTTAAAGGTACCTGAGG
AST1586 R	AGCAGAACAATTTGCTGAGGGC
AST-SEQ2R	GGTGAATATCCCTGCCTAACTCTAT
AST-SEQ2F	GGTTTAACATAACAAATTGGCTGTGGTATATAA
AST-SEQ3F	ATGGGTGGCAAGTGGTCAAA
Cloning Analysis Fwd	ACCTGCCAACCAAAGCGAGAAC
Cloning Analysis rev	TCAGGGTTATTGTCTCATGAGCG
Rev16	ATGGGAGGGGCATACATTGCT
Rev17	CCTGGAGCTGTTTAATGCCCCAGAC

To generate control transcripts, 1ug of plasmid was digested using restriction enzyme PmeI (New England Biolabs #R0560S) for *Ast* transcripts or ZraI (New England Biolabs #R0659S) for *Env* transcripts following the conditions outlined for these enzymes. The digested DNA was then purified using the QIAquick PCR Purification Kit (Qiagen # 28104) and concentrated using ethanol precipitation as described previously. RNA was transcribed using HiScribe T7 quick high yield RNA synthesis kit (New England Biolabs #E2050s) for *Ast* RNA or the HiScribe SP6 quick high yield RNA synthesis kit (New England Biolabs #E2070s) for *Env* RNA. Transcribed RNA was purified using a Monarch RNA cleanup kit (New England Biolabs #T2030) and quantified by nanodrop. The RNA was diluted to $1x10^6$ copies/mL and stored at -80°C.

To ensure complete digestion of plasmid DNA, approximately 10 thousand copies of the transcript were spiked into extracted nucleic acid from $1x10^5$ CEM cells and DNase treated as described above. The control transcripts were used for cDNA synthesis using random hexamer primers (Thermo Scientific #SO142) in the following reaction: 2.5µL of 10mM dNTPs and 2.5µL of 2µM random hexamer primer were added to each 20µL RNA

sample. The RNA was denatured at 65°C for 10 min, then immediately cooled at -20°C for 1 min. Next, 25μL of SuperScript III Reverse Transcriptase (catalog no. 18080-044; Invitrogen) master mix was added to the denatured RNA. Master mix was made by combining 10μL of 5X first strand buffer, 1μL of 0.1 M DTT, 12μL of RNase-free water, 1μL of 40U/μL RNaseOUT recombinant ribonuclease inhibitor (catalog no. 10777-019; Invitrogen), and 1μL of 200U/μL SuperScript III Reverse Transcriptase. cDNA was synthesized at 25°C for 5 minutes, 55°C for 1 hour, 70°C for 15 minutes, then cooled to 4°C. RNA was removed from the synthesized cDNA with the addition of 1μl RNase H (5 units) (NEB: Cat#M0297S) and incubation at 37°C for 20 minutes. Finally, enzymes were inactivated at 75°C for 10 minutes and the cDNA was cooled to 4°C and used immediately or stored at -80°C.

RESULTS

Development of the Ast CARD-SGS Assay

To detect expression of antisense transcripts in single HIV-1 infected cells, we adapted the CARD-SGS assay (Wiegand et al. 2017) to detect antisense transcripts throughout the provirus. Previously, the CARD-SGS assay developed by Wiegand, et al. (**Figure 7**) was used to characterize the levels and the percent of infected cells expressing unspliced HIV RNA in single infected cells from donors on ART by amplifying the *gag-pol* sub-genomic region (1.3-kb fragment of p6, protease and reverse transcriptase). Briefly, PBMCs from donors were divided into aliquots at an endpoint for infected cells with HIV RNA. HIV-1 RNA was extracted from each aliquot and cDNA was synthesized and spread across a 96-well microtiter plate for nested PCR amplification. PCR positive wells on each plate were sequenced by Sanger. The resulting sequences were analyzed using neighbor-joining distance trees. Identical sequences from the same aliquot were likely derived from a single infected cell, whereas RNA sequences from different aliquots arose from different cells. The CARD-SGS assay has been shown to be sensitive enough to detect a single cell expressing HIV-1 RNA in the P6-RT sub-genomic region.

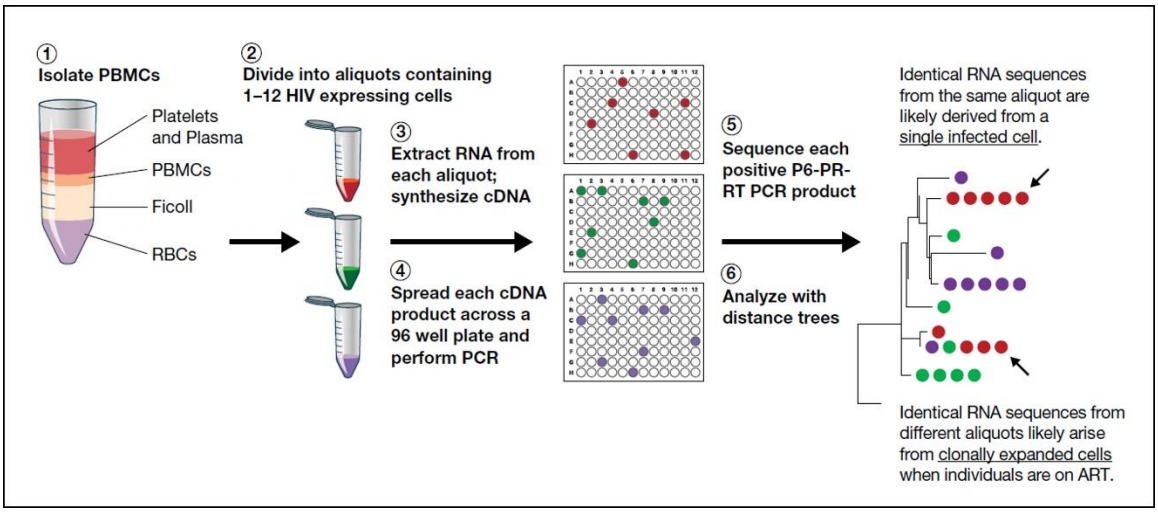


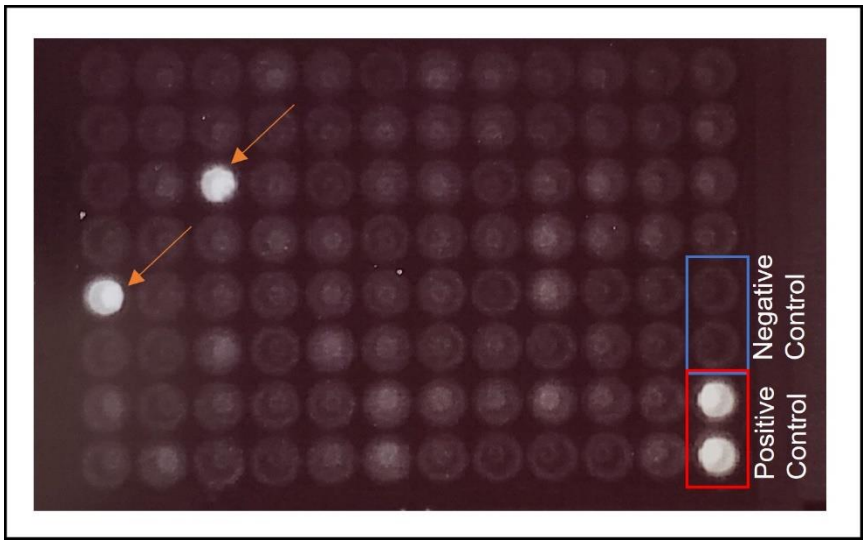
Figure 7: Workflow for CARD-SGS
Adapted from Wiegand et al. (2017) with permission.

To detect and quantify the fraction of cells expressing *Ast* RNA in donors on ART, primers were designed to synthesize cDNA and amplify and sequence the *Ast* region (nt7496-9237 in HXB2) (**Table 5**). To validate the *Ast* version of the CARD-SGS assay prior to evaluating donor samples, ACH2 cells, which are a chronically HIV-1 infected cell line and have been shown to contain *Ast* (Saayman et al. 2014; Zapata et al. 2017) were used. RNA was extracted from approximately 100 ACH2 cells spiked into $1x10^5$ uninfected CEM cells. Expression of antisense transcripts are about 2500x less abundant than sense transcripts (Zapata et al. 2017) and, Wiegand et al. (2017) found that 12-92 copies of unspliced sense HIV-1 RNA could be detected from 10 ACH2 cells using the CARD-SGS method. Therefore, we estimated that 100 ACH2 cells per aliquot was near endpoint and found that this dilution yielded, on average, 13/96 PCR wells positive for antisense HIV RNA.

Previous studies reported the potential of self-priming during cDNA synthesis due to the RNA secondary structure in *Ast* region, which can result in PCR amplification artifacts (Mancarella et al. 2019; Zapata et al. 2017). To investigate this possibility, cDNA

synthesis was performed excluding the cDNA primer. Positive PCR wells were observed and sequences were verified by Sanger (**Figure 8**), confirming self-priming of HIV RNA.

Figure 8: Self-Priming During Reverse Transcription



A cDNA reaction without a cDNA primer was performed using RNA extracted from 100 ACH2 cells in $1x10^5$ CEM cells. cDNA was PCR amplified using *pol* primers. Positives wells were detected and confirmed by Sanger sequencing. The negative control wells did not have template (blue box), and the positive control wells contained ACH2 DNA (red box).

To address this issue, a tagged RT primer as described in Zapata, et al. (Zapata et al. 2017) used in their highly sensitive, strand-specific quantitative RT-PCR assay, was tested for synthesis and amplification of *Ast* in our assay. The tagged RT primer contained an unrelated, exogenous sequence at the 5' end of the primer, while the 3' end anneals to the target RNA. This RT primer generated cDNA carrying the exogenous sequence at its 5' end, which served as the template for the forward primer during the first round of amplification (**Figure 9**). Thus, the tag ensures that only transcripts from antisense RNA were amplified from respective cDNA rather than self-primed templates. This was confirmed by performing cDNA synthesis with and without a cDNA primer on RNA from

an aliquot of 100 ACH2 cells in $1x10^5$ CEM cells. The tag sequence was then used as the forward primer in a nested PCR reaction of *Ast*. No positive wells were detected from cDNA synthesized without a cDNA primer; whereas positive wells were detected from cDNA synthesized with the tagged cDNA primer. A positive control in two reaction wells indicated that the PCR reaction was working even when no positive experimental wells were detected without the cDNA primer. Therefore, the tagged primer approach ensured detection of only antisense transcripts.

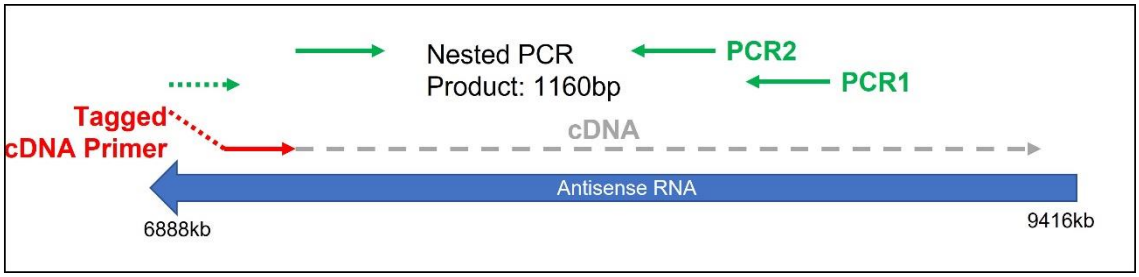


Figure 9: *Ast* Amplification Strategy
The tagged RT primer contained an unrelated, exogenous sequence at the 5' end, while the 3' end anneals to the target RNA. This RT primer generated cDNA carrying the exogenous sequence at its 5' end, which served as the template for the reverse primer to bind during PCR 1 amplification.

To confirm that DNA digestion after nucleic acid extraction protocol was complete, 100, 500, 1000 and 10,000 ACH2 cells were spiked into $1x10^5$ or $1x10^6$ uninfected CEM cells. Nucleic acid was isolated, DNA was digested, and a cDNA reaction was performed excluding the reverse transcriptase enzyme. The entire reaction was then spread across a 96-well plate and nested PCR was performed. We found that up to 500 infected ACH2 cells in a background of $1x10^5$ uninfected CEM had little to no DNA remaining. However, when greater than 500 infected ACH2 cells or greater than $1x10^6$ total cells were assayed, digestion of DNA was incomplete and the amount of DNA remaining was detected in the assay (**Figure 10**). To account for this possibility when assaying donor samples, a no RT

enzyme control was used whenever the number of infected cells per aliquot was greater than 500 and the total number of cells assayed per aliquot was more than $1x10^6$.

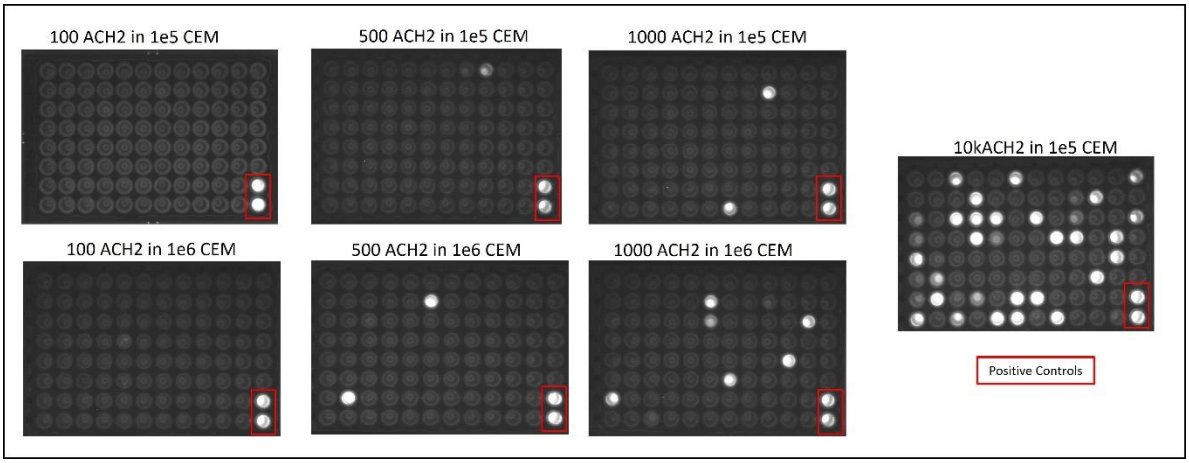


Figure 10: Confirmation of DNA Digestion During RNA Extraction
cDNA reactions were performed excluding the reverse transcriptase enzyme. The entire reaction was then spread across a 96 well plate and nested PCR was run with p6-RT primers. ACH2 DNA was added in 2 wells as a positive control (red box).

To improve sensitivity and efficiency of the assay, reverse transcription and nested PCR amplification concentrations and conditions were adjusted from those previously described (Wiegand, et al. 2017). The number of positive PCR reactions observed with the correct band size was used to assess various reaction conditions. For the cDNA synthesis, adjustments were made to the concentration of primer and RT enzyme, denaturation temperature and time, and annealing temperature and extension time (**Figure 11**). Nested PCR conditions were adjusted to recommended Platinum II DNA polymerase reaction conditions and to the length of the amplicon. Primers were specifically designed to anneal at 60ºC as required by the enzyme.

A

cDNA Condition Number	Amount of primer	Amount of Superscript	Denaturation Temp	Denaturation Time	Annealing Temp	Extension time	Number of positives
1	2.5µL of 2µM	0.5µL (100U)	65°C	10 minutes	55°C	50 minutes	1
2	2.5µL of 1µM	1µL (200U)	85°C	10 minutes	45°C	60 minutes	1
3	2.5µL of 1µM	1µL (200U)	95°C	5 minutes	45°C	60 minutes	1
4	2.5µL of 1µM	1µL (200U)	85°C	10 minutes	45°C	60 minutes	0
5	2.5µL of 2µM	1µL (200U)	85°C	10 minutes	45°C	60 minutes	6

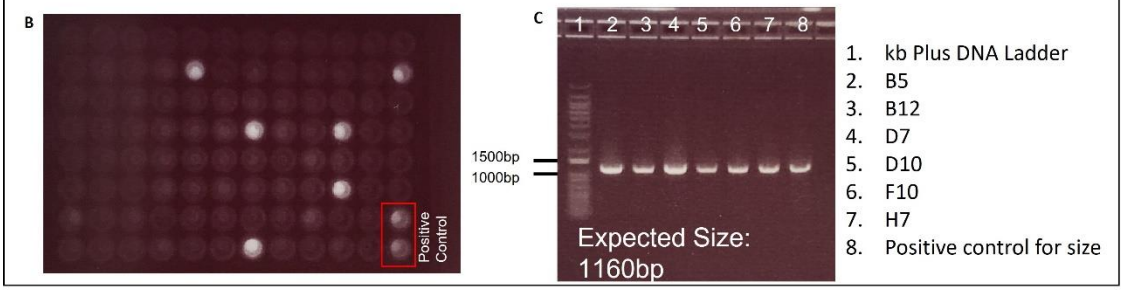
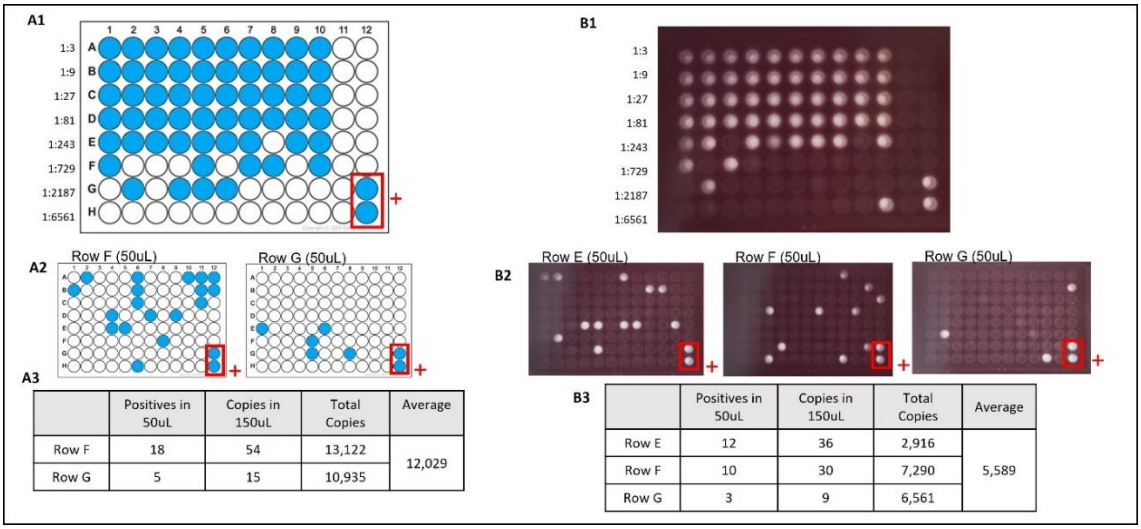


Figure 11: Optimizing *Ast* cDNA Synthesis
Various cDNA conditions were tested using RNA extracted from 100 ACH2 in $1x10^5$ CEM cells. A) Table of adjusted cDNA conditions and the number of positives generated. B) CARD-SGS detection plate from condition set 5. C) Positives from the detection plate run on an agarose gel.

To determine the sensitivity of the *Ast* CARD-SGS assay, in vitro transcribed *Ast* RNA molecules were generated using a pMiniT-AST vector with a T7 promoter and used as a standard. The concentration of RNA was measured by nanodrop and approximately 10,000 copies of the transcripts were spiked into nucleic acids extracted from $1x10^5$ CEM cells. First, the number of transcripts was quantified by synthesizing cDNA with random primers and performing endpoint PCR with primers and a 5' Hex labeled probe within the rev response element (RRE) of the envelope coding gene (nt7741-7847 in HXB2). Primers and probes were adapted from Bruner et al. to optimize them for the best signal to noise ratio (Bruner et al. 2019). After determining the starting number of transcripts, *Ast* CARD-SGS was performed on a second aliquot and the number of positives was compared to the known number of transcripts (**Figure 12**). These methods showed that the modified *Ast* CARD-SGS assay had a sensitivity of ~50%. The 50% sensitivity is likely due to the inclusion of a tag on the cDNA primer and to the secondary structure of the RNA. A digital

PCR assay, discussed later, was developed and used as an additional and more sensitive detection method to quantify the levels of *Ast* transcript expressed in infected cells from donors on ART.



	Positives in 50uL	Copies in 150uL	Total Copies	Average
Row F	18	54	13,122	12,029
Row G	5	15	10,935	

	Positives in 50uL	Copies in 150uL	Total Copies	Average
Row E	12	36	2,916	5,589
Row F	10	30	7,290	
Row G	3	9	6,561	

Figure 12: Determining *Ast* CARD-SGS Sensitivity
~10,000 copies of in vitro transcribed *Ast* RNA were spiked into $1x10^5$ CEM cells. A) Using random hexamer primers for cDNA synthesis and RRE primers and probes for qPCR, the number of transcripts were quantified per aliquot. A1) Dilution plate. 2µL of each cDNA dilution added to each PCR reaction well. A2) Plates of 50µL of select dilutions thought to be at endpoint. A3) Average total number of calculated copies based off the number of positives in the selected dilutions plates. B) The number of transcripts per aliquot were also quantified by *Ast* CARD-SGS. B1) Dilution plate. 2µL of each cDNA dilution added to each reaction well. B2) Plates of 50µL of select dilutions thought to be at endpoint. B3) Average total number of calculated copies based off the number of positives in the selected dilutions plates.

Adaption of the CARD-SGS Assay for Other Antisense Regions

The CARD-SGS assay was also modified to detect and quantify antisense transcripts in *gag* (nt764-2281 in HXB2) (Sanders-Buell et al. 1995) and in *pol* (nt1849-3500 in HXB2) (Wiegand et al. 2017) and (nt3996-5270 in HXB2) regions using cDNA and nested PCR primers described in Table 5. The same exogenous tag as described above was added to the cDNA synthesis primer and the "tag" portion acted as the target for the PCR1 forward primer (Tag Fwd). Thermocycling conditions were modified based on the size of the amplicon. These additional antisense assays were tested on RNA extracted from aliquots of 100 ACH2 cells spiked into $1x10^5$ uninfected CEM cells. Positive wells were observed and sequenced to confirm amplification of antisense transcript from the three additional regions of the provirus (**Figure 13**).

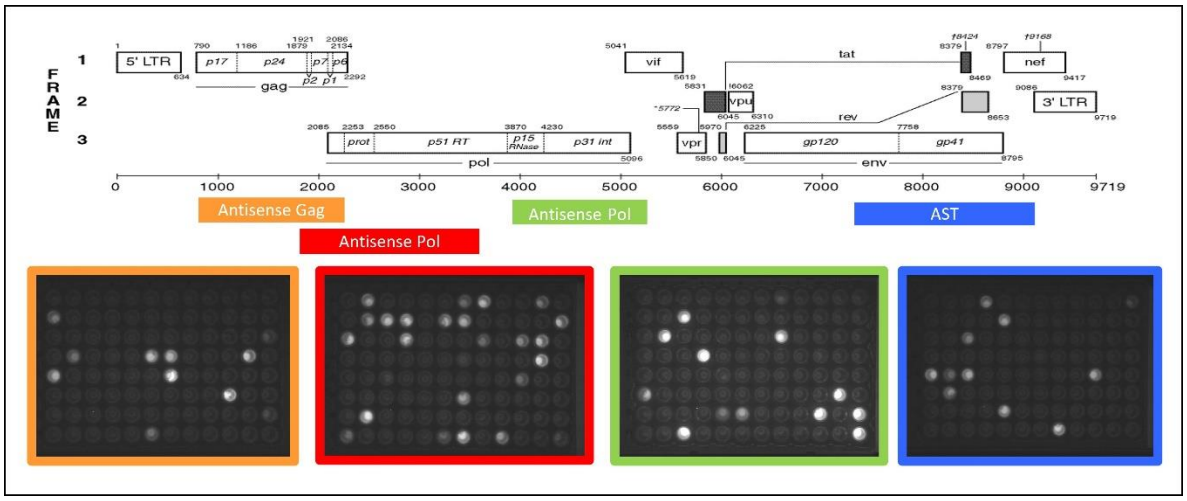


Figure 13: Amplification and Detection of Antisense Transcripts in *gag* and *pol* (P6-RT and integrase regions) and *Ast* by CARD-SGS.
The CARD-SGS assay was modified using the tagged primer approach to synthesize cDNA and to amplify antisense transcripts. In each assay, RNA from 100 ACH2 cells in $1x10^5$ CEM cells per aliquot was tested.

Development of a Digital PCR Assay with Primers and Probes

To determine the levels of *Ast* and *env* RNA with a more sensitive method, a digital PCR assay was developed. Compared to *Ast* CARD-SGS, the limitations of this assay are that it cannot determine the number of *Ast* or *Env* RNA from a single cell and the positive wells cannot be sequenced. However, this assay measured the level of *Ast* and *env* RNA in to compare the level of sense and antisense transcription.

The primers and probes used to detect *Ast* transcripts were adapted from the highly sensitive, strand-specific quantitative RT-PCR assay previously described (Zapata et al. 2017). Additionally, this assay uses the same cDNA primer as the *Ast* CARD-SGS assay. Envelope primers were developed within the same region as the *Ast* primers, however, the forward primer was shifted an additional 141bp upstream due to poor performance of the original forward primer located at a position similar to the *Ast* assay.

To determine the sensitivity of both digital PCR assays, in vitro transcribed *Ast* and *env* RNA molecules were generated. As described previously, the concentration of RNA was measured by nanodrop and ~10,000 copies of the transcript were spiked into $1x10^5$ CEM cells. The number of transcripts was quantified by synthesizing cDNA using random hexamer primers and completing endpoint PCR with primers and probes that bind the RRE region of the transcript (Bruner et al. 2019). After determining the number of transcripts per aliquot, *Ast* and *env* digital PCR was performed, and the number of positives was compared to the known number of transcripts. These methods showed that both assays had a sensitivity of ~100% (**Table 10**).

Table 10: Digital PCR Assay Sensitivity

Transcript	cDNA Primer	Total Copies by Digital PCR	Average Number of Copies	Sensitivity
Ast Transcript	Random Hexamer Primers	81	108	~100%
		135		
	AST-Tag RT Primer	189	176	
		162		
Env Transcript	Random Hexamer Primers	1600	1600	~100%
	ENV-Tag RT Primer	1800	1800	

HIV-1 Antisense Expression in Donors on ART

All donors included in this study were from the SCOPE cohort and were on long-term suppressive ART (**Table 2**). PBMCs were obtained from three donors (participant identifiers (PIDs) 2669, 1079, and 1683), and enriched CD4+ T cells and HLA-DR- (Resting) CD4+ T cells were obtained for one donor (3162). One timepoint during continuous suppression for a median of 5.4 years (5.4-12.8 years) was obtained for three of the donors (1079, 1683, and 3162). Samples from multiple timepoints preceding and following a recent ART interruption were obtained for one donor (2669).

PBMCs were divided into 8-10 aliquots of $1x10^6$ cells for examination by CARD-SGS for *Ast*. One aliquot of PBMCs was used to extract DNA for each donor and timepoint to quantify the number of infected cells per million PBMCs as described in the methods section. Enriched total CD4+ T cells and HLA-DR- (Resting) CD4+ T cells were received as cell pellets in aliquots of 20k and $1x10^6$ cells, respectively. DNA was also extracted

from one of each to measure the number of infected cells. RNA from cell aliquots were also assayed by the digital PCR assay.

Extracted DNA from each donor was also used to obtain sequence information to develop donor specific primers. CARD-SGS was performed on extracted DNA using primers that are outside of the region amplified by the *Ast* CARD-SGS assay to ensure primer matching (**Table 5**). The resulting amplicons were sequenced by Sanger and were used to design donor specific primers. *Ast* CARD-SGS and digital PCR primers as well as *env* digital PCR primers were modified for specificity for each donor as needed (**Table 11,** patient-specific mutations shown in red).

Table 11: Donor Specific Primers

Primer Name	Primer Sequence (5'-…-3')
2669_AST IF	AGCAGAACAATTTGCTGAGRGC
2669_AST IR	GTCATTGGTCTTAAAGGTACCTGRGG
2669 AST Rev-qPCR	TGATGAACATCTAATTAGTCCTTTGA
2669_ENV-Tagged Primer	CTGATCTAGAGGTACCGGATCCTGATGAACATCTAATTAGTCCTTTGA
1079_AST Rev-qPCR	TGATGAACATCTAATTTSTCCMCTGA
1079_ENV-Tagged Primer	CTGATCTAGAGGTACCGGATCCACATCTAATTTSTCCMCTGA
1683_AST-Tagged Primer	CTGATCTAGAGGTACCGGATCCAACATGTGGCAAGAAATAGG
1683_AST OR	TGGTACTAGCTTGTAGCACCACCC
1683_AST IF	AGCAGAACAATTTGCTGAGAGC
1683_AST Rev-qPCR	TGATGAACATCTAATCAGTCCGCTGA
1683_ENV-Tagged Primer	CTGATCTAGAGGTACCGGATCCACATCTAATCAGTCCGCTGA
1683_ENV Fwd-qPCR	ACAAATTATAAACATGTGGCAAGAAATAGG
3162_AST-Tagged Primer	CTGATCTAGAGGTACCGGATCCAACATGTGGCAGGAAGTAGG
3162_AST OR	TGGTACTAGCTTGTAGCACCAYCC
3162_AST IF	AGCAGAACAATTTGCTGAGRGC
3162_AST Rev-qPCR	TGATGAACAGTTAATGAGTCCCCTGA
3162_ENV-Tagged Primer	CTGATCTAGAGGTACCGGATCCACAGTTAATGAGTCCCCTGA
3162_ENV Fwd-qPCR	ACAAATTRTAAACATGTGGCAGGAAGTAGG

RNA was extracted from eight to ten aliquots of $1x10^6$ donor PBMCs and used for *Ast* CARD-SGS with donor specific primers. Only two aliquots of PID 3162 resting cells could be evaluated because of sample availability. A median of ~4% (range: 1.1 - 5.6%) of infected PBMCs contained the 1.7kb fragment of *Ast* RNA measured by this method. The level of expression per cell averaged one copy but ranged from 1-3 copies except for one donor sample from PID 1079 which was found to have one cell with 30 copies of the 1.7kb *Ast* transcript (**Table 12**). For donor 2669, *Ast* expression levels over the four timepoints preceding and following a recent ART interruption did not show any statistically significant differences in the levels or fraction of cells with *Ast* by a one-way ANOVA (p=0.24) (**Figure 14**). Analysis of CD4+ T cells and HLA-DR- (resting) T cells revealed that the expression of *Ast* in resting cells is very low with only 0.01% of infected cells with the antisense transcript. However, in total CD4+ T cells, *Ast* expression was 600 times higher (0.01% vs 6.3% of infected cells with *Ast* RNA). This result suggests that the majority of the *Ast* transcripts detected in total CD4+ T cells may be expressed from HLA-DR+ (Activated) CD4+ Memory T cells, although studies on additional donor samples are needed to confirm this finding.

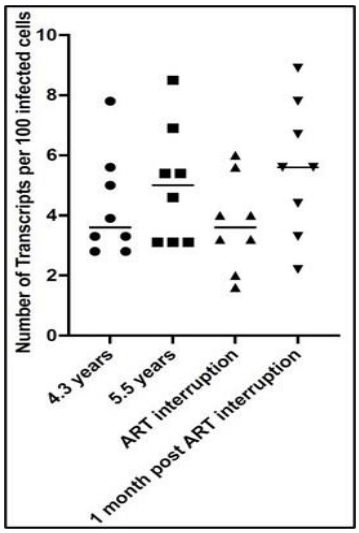


Figure 14: No Significant Change in the Fraction of Cells Expressing *Ast* Over Time or During ART Interruption in PBMCs from PID 2669

Table 12: *Ast* CARD-SGS Results

PID	# infected cells with *env* DNA per million PBMCs	Estimated # infected cells assayed for HIV RNA	# HIV *Ast* RNA sequences obtained	Estimated # of infected cells with *Ast* RNA*	Estimated % of infected cells with HIV *Ast* RNA	Average *Ast* RNA copies per cell (Range)
2669 (T1)	180	1440	67	63	4.4%	1.1 (1-3)
2669 (T2)	130	1040	54	52	5%	1.0 (1-2)
2669 (T3)	250	2000	86	74	3.7%	1.2 (1-3)
2669 (T4)	90	720	41	40	5.6%	1.0 (1-2)
1079	80	720	43	12	1.7%	3.6 (1-30)
1683	188	1880	22	20	1.1%	1.1 (1-2)
3162 (eCD4+)	3100	496	34	31	6.3%	1.1 (1-2)
3162 (Resting)	2700	4050	40	37	0.01%	1.0 (1-2)

Neighbor joining trees were constructed from cell-associated *Ast* RNA sequences and rooted to an HIV-1 subtype B consensus sequence (**Figure 15 and 16**). Each square on the tree represents one RNA molecule, and the colors represent the aliquot from which they came. If identical sequences are from the same aliquot, they are considered to have originated from a single cell. Visualization of the sequences in this manner allows for the detection of "high-expressing cells", like the one with 30 copies of *Ast* in PID 1079. The median diversity measured by average pairwise distance (APD) was 3.2% with a range of 1.2 - 4.2%. The APD was also calculated for each donor without the inclusion of the hypermutants, which had a median of 2.5% and a range of 0.6%-3.4%. Hypermutants are typically the result of mutation by host restriction factors, such as APOBEC3G or 3F.

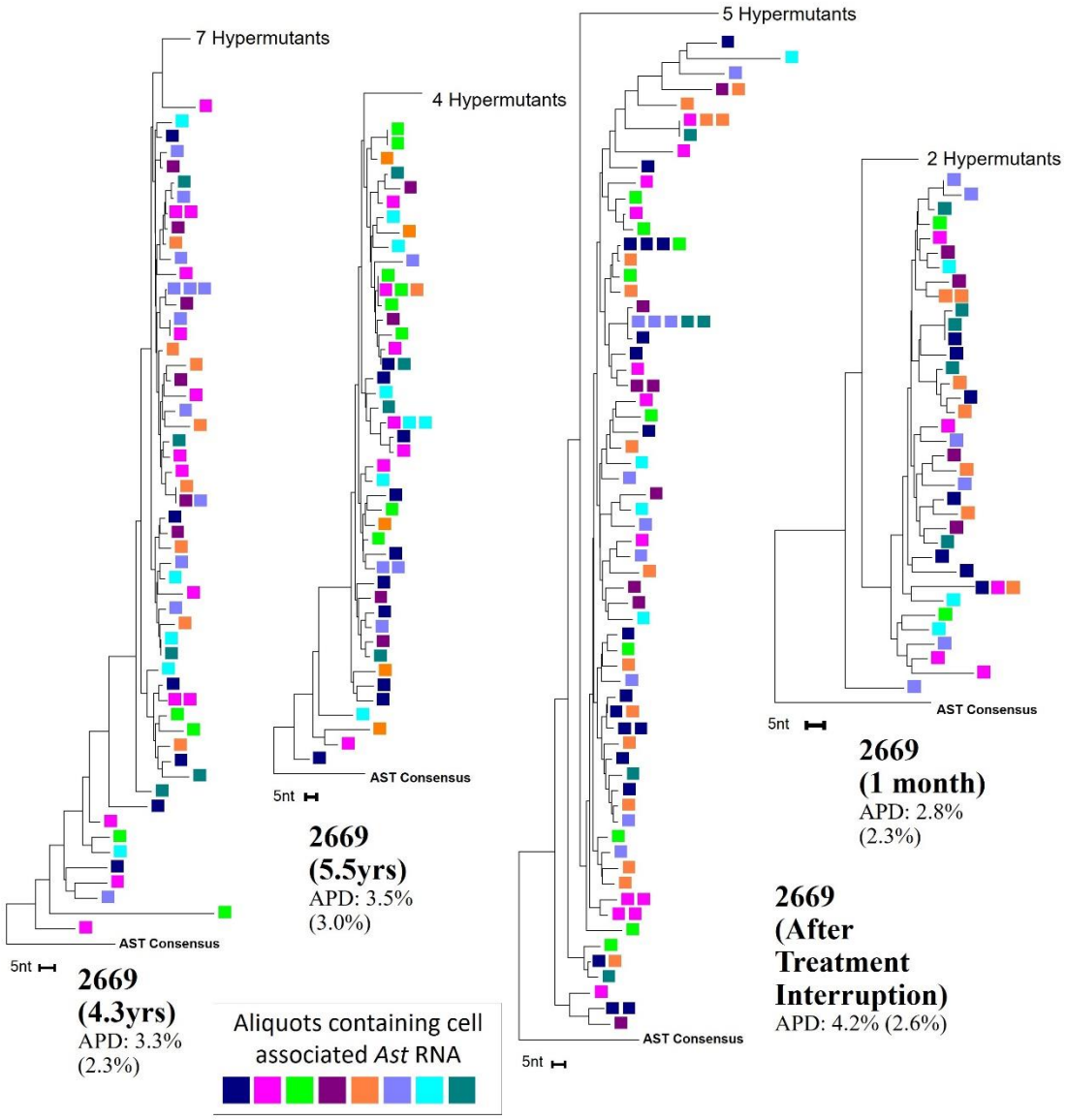


Figure 15: **PID 2669 Neighbor Joining Trees.**
Phylogenetic trees were constructed from cell-associated *Ast* RNA from different timepoints from PID 2669 (4.3 years on ART, 5.5 years on ART, during treatment interruption, an 1 month after resuming ART). Trees are rooted on an HIV-1 subtype B consensus sequence. Average pairwise distance is expressed as: APD (APD without hypermutants).

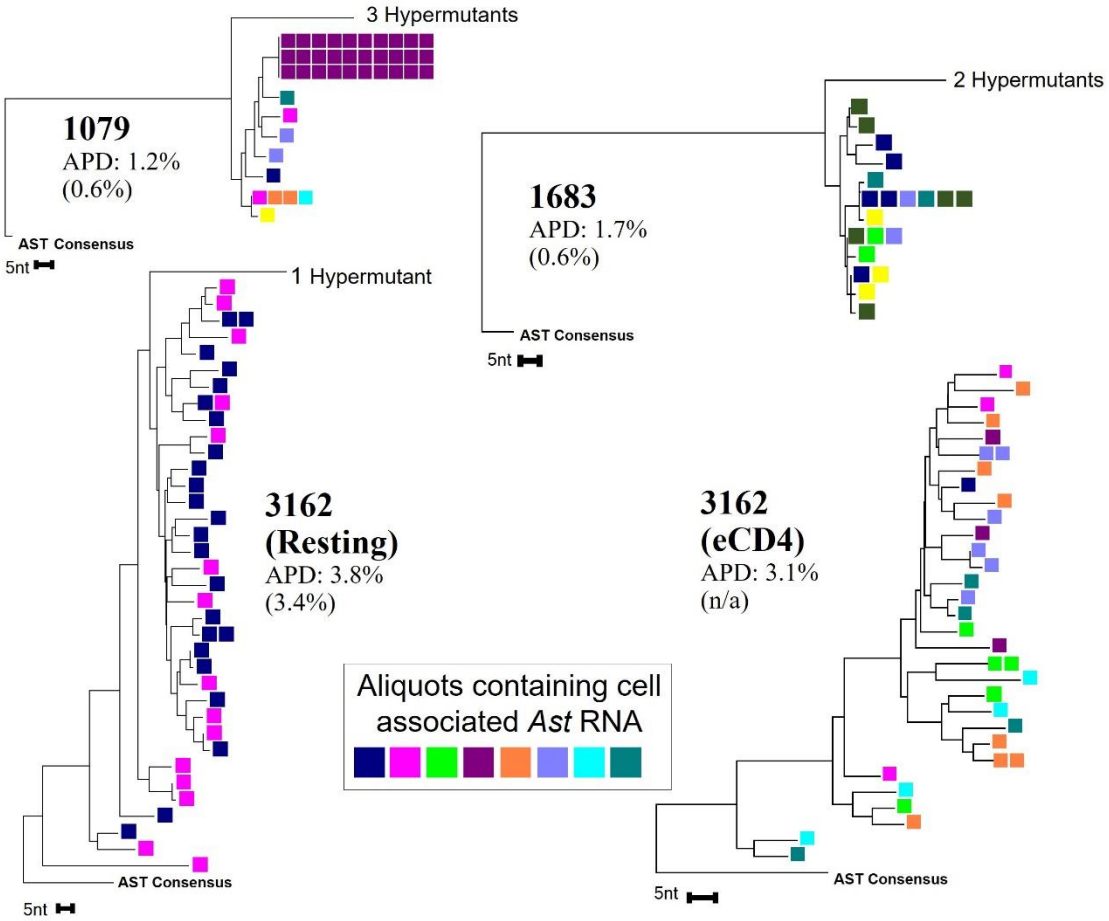


Figure 16: **3162, 1079, and 1683 Neighbor Joining Trees.**
Trees were constructed from cell-associated *Ast* RNA and rooted to a HIV-1 subtype B consensus sequence. Average pairwise distance is expressed as: APD (APD without hypermutants).

Ast CARD-SGS results were compared to the levels of expression of the HIV *pol* gene in the same donors (McManus et al. 2019). We found no significant differences in the fraction of cells with antisense and sense HIV transcripts (**Figure 17A**) or in the levels of their expression in single infected cells (**Figure 17B**).

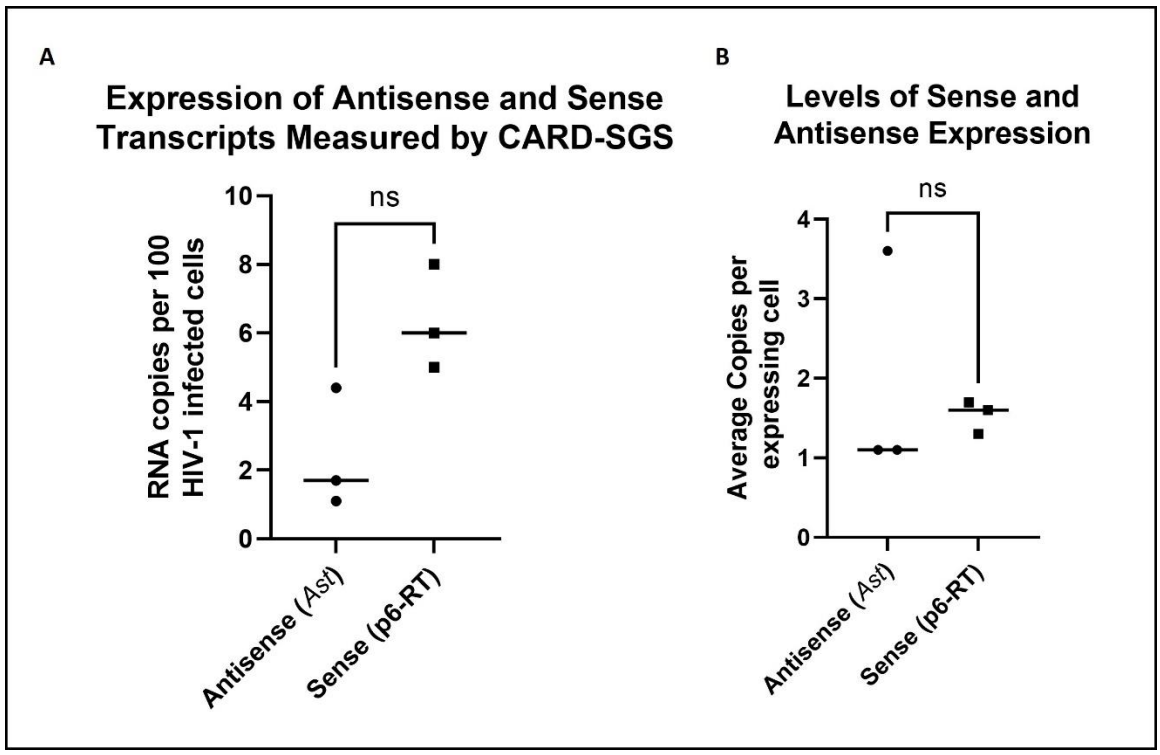


Figure 17: Sense vs. Antisense CARD-SGS Results
A) Comparison of the percent of infected cells with *Ast* and sense *pol* transcripts. B) Comparison of the levels of *Ast* and sense *pol* transcripts in single infected cells.

Results from the digital PCR assay developed to evaluate the expression of sense and antisense transcripts from the same region of envelope also indicated that there was no significant difference in the levels of sense and antisense expression in the donors sampled. A median of 26 copies per 100 infected cells (range 11-69) had *Ast* RNA, while a median of 19 copies per 100 infected cells (range 6-35) had *env* sense RNA (**Figure 18**).

The expression of *Ast* RNA from HIV-1 infected cells measured by the digital PCR assay was much higher than those measured by the CARD-SGS assay. This difference is likely due to the different sensitivities of the assays (CARD-SGS assay has a sensitivity of ~50%, while the digital PCR assay has a sensitivity closer to 100%), and the different length of the RNA molecules being detected. If the RNA has been degraded, it is more likely that an assay detecting a shorter piece of RNA will be successful.

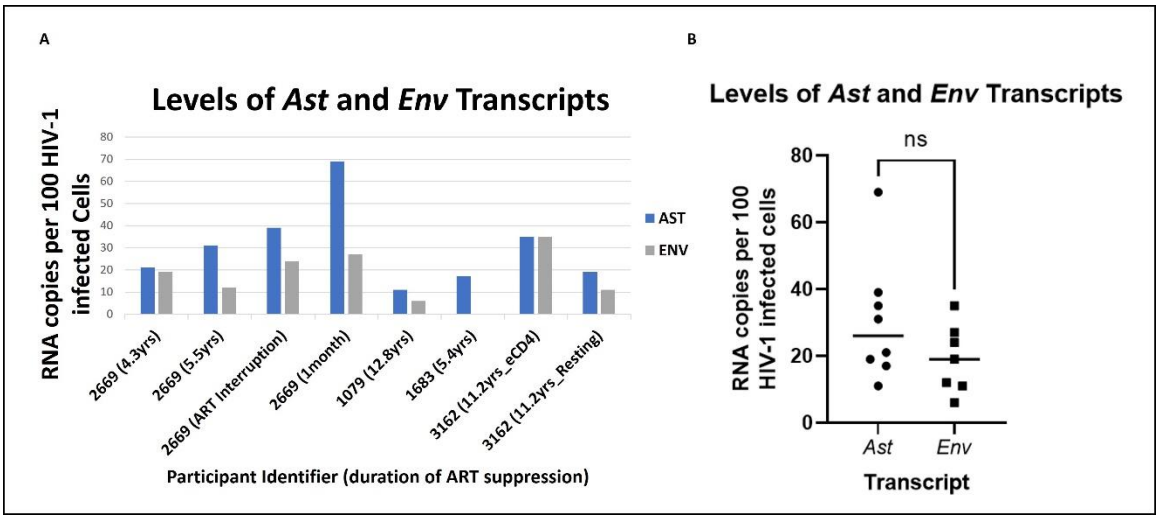


Figure 18: Digital PCR Results
A) Levels of *Ast* and *Env* transcripts detected by digital PCR for each donor and timepoint. B) A comparison of the levels of *Ast* and *Env* transcript expression.

RNA extracted from donor 2669, sample T4 (1 month on ART) was also assayed for the presence of antisense transcripts in *gag* and *pol*, as well as the *Ast* genomic region (**Figure 19**). Positive PCR wells targeted various regions of the HIV genome were observed and sequenced. Sequences matched antisense HIV confirming that antisense transcription could include the full-length HIV.

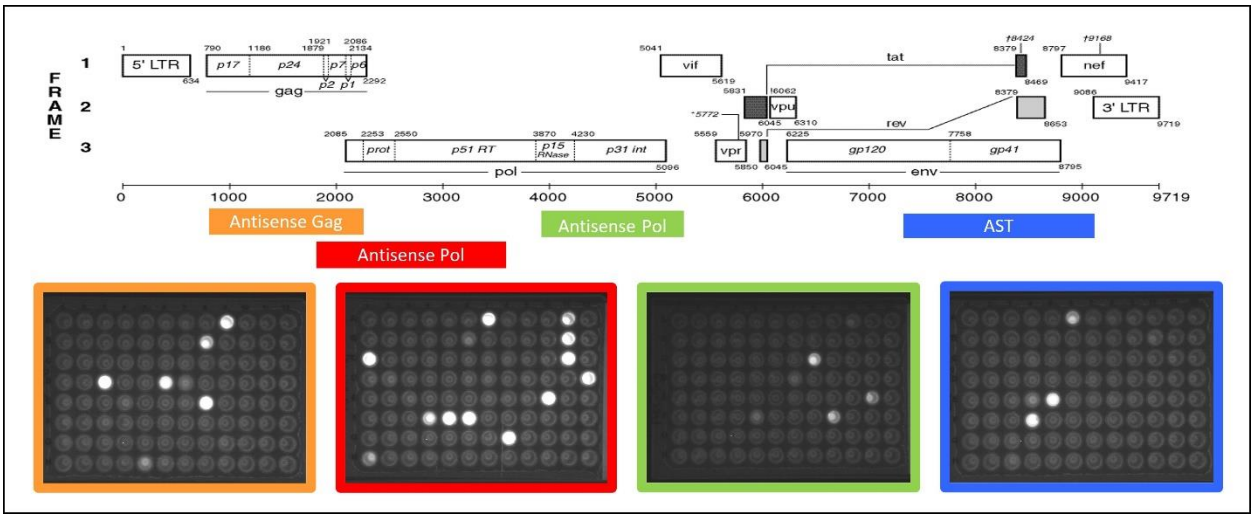


Figure 19: Detection of Antisense Transcript Across the Proviral Genome
The modified CARD-SGS assay using tagged primers to amplify antisense transcripts in the *gag*, *pol* (P6-RT), and *pol* (integrase) regions was run on RNA extracted from $1x10^6$ PBMC from donor 2669 sample from 1 month suppressed on ART (T4).

DISCUSSION

Although HIV-1 *Ast* RNA has been shown to inhibit viral replication and promote the establishment and maintenance of latency in vitro (Saayman et al. 2014; Zapata et al. 2017), studies investigating *Ast* expression in vivo have been limited. Therefore, we sought to investigate if *Ast* is expressed in single HIV-1 infected cells from donors on ART to understand how the expression of *Ast* in vivo may contribute to HIV-1 persistence in PLWH. The CARD-SGS assay was adapted to measure antisense transcripts in single cells and a digital PCR approach was developed with increased sensitivity to measure antisense and sense transcripts from donors on long-term ART. Using these new assays, we determined the levels of antisense transcription and the fraction of HIV-1 infected cells with *Ast* RNA. Additionally, we compared the levels of *Ast* to the expression of other HIV-1 genes. We found that antisense transcripts were present in *gag*, *pol*, and *env*, suggesting that antisense transcription may include the entire length of the HIV-1 provirus. Finally, we examined the levels and fraction of HIV-1 infected cells with *Ast* in donors on ART and during antiretroviral therapy interruption (ATI).

We found that a median of ~5% of infected PBMCs contained the 1.7kb fragment of *Ast* RNA measured by CARD-SGS and those levels of expression ranged from 1-30 copies per cell. These fractions and levels are similar to those observed previously for sense HIV-1 transcripts in the same donors. Using the strand-specific digital RT-PCR approach, which detects a shorter fragment of *Ast* RNA within the same region, we measured a median of 26 copies per 100 infected PBMCs. A similar fraction of infected cells contained sense *pol* and *env* RNA in most donor samples.

These findings confirm expression of HIV-1 *Ast* RNA in donors on ART and indicate that antisense transcripts may span the entire length of the HIV-1 provirus, consistent with in vitro studies. Detection of *Ast* RNA in unstimulated cells from donors on ART warrants the investigation of its bifunctional role as a coding and a regulatory RNA in vivo. Targeting *Ast* expression may be an HIV-specific approach to reversing or inducing viral latency in vivo.

The fact that the digital PCR assay that detects a shorter fragment of *Ast* RNA found a higher number of transcripts is likely informative about the stability of the RNA. If the RNA is being rapidly degraded, then it would be likely that this assay could still detect it. Rapid degradation of this antisense transcript would mean that its expression could be easily and quickly turned on or off, and that detecting only one molecule would not mean that expression levels are low. Rather it could mean that expression may be high but that the RNA is degraded as soon as it is transcribed. Additionally, it could also suggest that there could be *Ast* RNA in all cells but that it is only detected in about 30% at a time.

Although these studies demonstrate active antisense HIV-1 transcription in vivo, there are still many open questions. For example, do the HIV-1 antisense transcripts detected in vivo influence HIV sense transcription? In other words, do they induce HIV-1 latency and, therefore, contribute to the persistence of HIV-1 during ART? How does expression of *Ast* differ in people in acute vs chronic infection, on ART vs not on ART, or in non-controllers vs elite controllers? How do levels of *Ast* RNA vary in different tissues and different cell subsets? What is the stability of Ast RNA and how does the stability (or instability) relate to its function in vivo? What fraction of *Ast* RNA molecules are read-through transcripts and do they have a role in HIV-1 latency?

Another open question is whether there are other antisense genes and/or antisense splice variants encoded in the HIV-1 genome. Several in vitro studies have described antisense transcripts that differ in start site, length, and polyadenylation status (Kobayashi-Ishihara et al. 2018; Landry et al. 2007; Michael et al. 1994). Our study provides preliminary insight into the length of the antisense transcript in vivo. We found that antisense HIV-1 transcripts are detectable in *gag*, *pol*, and *env*. However, we do not yet know if these RNA molecules constitute one continuous transcript or multiple antisense transcripts. Future studies will investigate the location of the *Ast* start site and the polyadenylation status of these transcripts in vivo.

Although in vitro studies have shown that antisense expression is driven by an NSP in the U3 region of the 3' LTR, the role of chromatin remodeling and epigenetics on the activity of the promoter is unknown. Preliminary studies in vitro have suggested that the 3' LTR NSP is under epigenetic control, and that latency reversing agents (LRAs) can increase antisense transcription from the 3' LTR (Li et al. 2021). These results potentially explain why LRAs have limited efficacy in "shock and kill" strategies, because simultaneous upregulation of antisense transcripts could counteract the effect of LRAs on sense transcription. An additional question to be answered is whether expression of both antisense and sense transcripts can occur in the same single cell in vivo. If the HIV-1 antisense transcript is acting as a lncRNA to promote epigenetic silencing of the HIV-1 5'LTR, cells expressing the antisense transcript should not produce sense transcripts. Future studies measuring levels of both sense and antisense transcription in the same cell could support the proposed mechanism of epigenetic silencing.

Altogether, the data generated here and the questions to be answered in the future may lead to future strategies that overexpress *Ast* to achieve successful "block and lock" of HIV proviruses, delaying or preventing viral rebound if ART is interrupted.

REFERENCES

Affram Y, Zapata JC, Gholizadeh Z, Tolbert WD, Zhou W, Iglesias-Ussel MD, Pazgier M, Ray K, Latinovic OS, Romerio F. 2019. The HIV-1 antisense protein ASP is a transmembrane protein of the cell surface and an integral protein of the viral envelope. J Virol. 93(21).

Archin NM, Keedy KS, Espeseth A, Dang H, Hazuda DJ, Margolis DM. 2009. Expression of latent human immunodeficiency type 1 is induced by novel and selective histone deacetylase inhibitors. AIDS. 23(14):1799-1806.

Arts EJ, Hazuda DJ. 2012. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med. 2(4):a007161.

Bangham CRM, Miura M, Kulkarni A, Matsuoka M. 2019. Regulation of latency in the human t cell leukemia virus, HTLV-1. Annu Rev Virol. 6(1):365-385.

Bentley K, Deacon N, Sonza S, Zeichner S, Churchill M. 2004. Mutational analysis of the HIV-1 LTR as a promoter of negative sense transcription. Arch Virol. 149(12):2277-2294.

Bet A, Maze EA, Bansal A, Sterrett S, Gross A, Graff-Dubois S, Samri A, Guihot A, Katlama C, Theodorou I et al. 2015. The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. Retrovirology. 12:15.

Billman MR, Rueda D, Bangham CRM. 2017. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome Open Res. 2:87.

Boltz VF, Ceriani C, Rausch JW, Shao W, Bale MJ, Keele BF, Hoh R, Milush JM, Deeks SG, Maldarelli F et al. 2021. Cpg methylation profiles of HIV-1 pro-viral DNA in individuals on art. Viruses. 13(5).

Briquet S, Richardson J, Vanhee-Brossollet C, Vaquero C. 2001. Natural antisense transcripts are detected in different cell lines and tissues of cats infected with feline immunodeficiency virus. Gene. 267(2):157-164.

Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, Fray EJ, Beg SA, Antar AAR, Jenike KM et al. 2019. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. Nature. 566(7742):120-125.

Bukrinsky MI, Etkin AF. 1990. Plus strand of the HIV provirus DNA is expressed at early stages of infection. AIDS Res Hum Retroviruses. 6(4):425-426.

Cassan E, Arigon-Chifolleau AM, Mesnard JM, Gross A, Gascuel O. 2016. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. Proc Natl Acad Sci U S A. 113(41):11537-11542.

Cavanagh MH, Landry S, Audet B, Arpin-Andre C, Hivin P, Pare ME, Thete J, Wattel E, Marriott SJ, Mesnard JM et al. 2006. HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated. Retrovirology. 3:15.

Chawla A, Wang C, Patton C, Murray M, Punekar Y, de Ruiter A, Steinhart C. 2018. A review of long-term toxicity of antiretroviral treatment regimens and implications for an aging population. Infect Dis Ther. 7(2):183-195.

Clerc I, Polakowski N, Andre-Arpin C, Cook P, Barbeau B, Mesnard JM, Lemasson I. 2008. An interaction between the human t cell leukemia virus type 1 basic leucine zipper factor (hbz) and the kix domain of p300/cbp contributes to the down-

regulation of tax-dependent viral transcription by hbz. J Biol Chem. 283(35):23903-23913.
Deeks SG, Overbaugh J, Phillips A, Buchbinder S. 2015. Hiv infection. Nat Rev Dis Primers. 1:15035.
Depledge DP, Ouwendijk WJD, Sadaoka T, Braspenning SE, Mori Y, Cohrs RJ, Verjans G, Breuer J. 2018. A spliced latency-associated VZV transcript maps antisense to the viral transactivator gene 61. Nat Commun. 9(1):1167.
Dufour C, Gantner P, Fromentin R, Chomont N. 2020. The multifaceted nature of HIV latency. J Clin Invest. 130(7):3381-3390.
Durkin K, Rosewick N, Artesi M, Hahaut V, Griebel P, Arsic N, Burny A, Georges M, Van den Broeke A. 2016. Characterization of novel bovine leukemia virus (BLV) antisense transcripts by deep sequencing reveals constitutive expression in tumors and transcriptional interaction with viral microrna. Retrovirology. 13(1):33.
Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ et al. 2019. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J Clin Invest. 129(3):988-998.
Fauci AS, Macher AM, Longo DL, Lane HC, Rook AH, Masur H, Gelmann EP. 1984. Nih conference. Acquired immunodeficiency syndrome: Epidemiologic, clinical, immunologic, and therapeutic considerations. Ann Intern Med. 100(1):92-106.
Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, Smith K, Lisziewicz J, Lori F, Flexner C et al. 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med. 5(5):512-517.
Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R et al. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science. 278(5341):1295-1300.
Friedman J, Cho WK, Chu CK, Keedy KS, Archin NM, Margolis DM, Karn J. 2011. Epigenetic silencing of HIV-1 by the histone h3 lysine 27 methyltransferase enhancer of zeste 2. J Virol. 85(17):9078-9089.
Gazon H, Lemasson I, Polakowski N, Cesaire R, Matsuoka M, Barbeau B, Mesnard JM, Peloponese JM, Jr. 2012. Human t-cell leukemia virus type 1 (HTLV-1) bzip factor requires cellular transcription factor jund to upregulate htlv-1 antisense transcription from the 3' long terminal repeat. J Virol. 86(17):9070-9078.
Giordani NV, Neumann DM, Kwiatkowski DL, Bhattacharjee PS, McAnany PK, Hill JM, Bloom DC. 2008. During herpes simplex virus type 1 infection of rabbits, the ability to express the latency-associated transcript increases latent-phase transcription of lytic genes. J Virol. 82(12):6056-6060.
Guidelines for the use of antiretroviral agents in adults and adolescents living with hiv. 2019. Clinical Info HIV.gov; [accessed]. https://clinicalinfo.hiv.gov/en/guidelines/adult-and-adolescent-arv/adverse-effects-antiretroviral-agents.
Kauder SE, Bosque A, Lindqvist A, Planelles V, Verdin E. 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog. 5(6):e1000495.

Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, Rehm C, Poethke C, Kovacs N, Mellors JW et al. 2014. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. PLoS Pathog. 10(3):e1004010.
Khorkova O, Myers AJ, Hsiao J, Wahlestedt C. 2014. Natural antisense transcripts. Hum Mol Genet. 23(R1):R54-63.
Kleinpeter AB, Freed EO. 2020. HIV-1 maturation: Lessons learned from inhibitors. Viruses. 12(9).
Kobayashi-Ishihara M, Terahara K, Martinez JP, Yamagishi M, Iwabuchi R, Brander C, Ato M, Watanabe T, Meyerhans A, Tsunetsugu-Yokota Y. 2018. Hiv ltr-driven antisense RNA by itself has regulatory function and may curtail virus reactivation from latency. Front Microbiol. 9:1066.
Kumar S, Stecher G, Tamura K. 2016. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870-1874.
Landry S, Halin M, Lefort S, Audet B, Vaquero C, Mesnard JM, Barbeau B. 2007. Detection, characterization and regulation of antisense transcripts in HIV-1. Retrovirology. 4:71.
Lee E, Bacchetti P, Milush J, Shao W, Boritz E, Douek D, Fromentin R, Liegler T, Hoh R, Deeks SG et al. 2019. Memory CD4 + T-cells expressing HLA-DR contribute to hiv persistence during prolonged antiretroviral therapy. Front Microbiol. 10:2214.
Levine AM. 1987. Non-hodgkin's lymphomas and other malignancies in the acquired immune deficiency syndrome. Semin Oncol. 14(2 Suppl 3):34-39.
Li R, Sklutuis R, Groebner JL, Romerio F. 2021. HIV-1 natural antisense transcription and its role in viral persistence. Viruses. 13(5).
Liu B, Zhao X, Shen W, Kong X. 2015. Evidence for the antisense transcription in the proviral r29-127 strain of bovine immunodeficiency virus. Virol Sin. 30(3):224-227.
Liu Z, Torresilla C, Xiao Y, Nguyen PT, Cate C, Barbosa K, Rassart E, Cen S, Bourgault S, Barbeau B. 2019. HIV-1 antisense protein of different clades induces autophagy and associates with the autophagy factor p62. J Virol. 93(2).
Longo DL, Steis RG, Lane HC, Lotze MT, Rosenberg SA, Preble O, Masur H, Rook AH, Fauci AS, Jacob J et al. 1984. Malignancies in the aids patient: Natural history, treatment strategies, and preliminary results. Ann N Y Acad Sci. 437:421-430.
Lusic M, Marcello A, Cereseto A, Giacca M. 2003. Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. EMBO J. 22(24):6550-6561.
Ma G, Yasunaga J, Matsuoka M. 2016. Multifaceted functions and roles of hbz in HTLV-1 pathogenesis. Retrovirology. 13:16.
Ma G, Yasunaga JI, Shimura K, Takemoto K, Watanabe M, Amano M, Nakata H, Liu B, Zuo X, Matsuoka M. 2021. Human retroviral antisense mRNAs are retained in the nuclei of infected cells for viral persistence. Proc Natl Acad Sci U S A. 118(17).
Maeda M, Shimizu A, Ikuta K, Okamoto H, Kashihara M, Uchiyama T, Honjo T, Yodoi J. 1985. Origin of human t-lymphotrophic virus i-positive t cell lines in adult t cell leukemia. Analysis of T cell receptor gene rearrangement. J Exp Med. 162(6):2169-2174.

Majerciak V, Yang W, Zheng J, Zhu J, Zheng ZM. 2019. A genome-wide epstein-barr virus polyadenylation map and its antisense RNA to EBNA. J Virol. 93(2).
Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF et al. 2014. Hiv latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science. 345(6193):179-183.
Mancarella A, Procopio FA, Achsel T, Crignis ED, Foley BT, Corradin G, Bagni C, Pantaleo G, Graziosi C. 2019. Detection of antisense protein (Asp) RNA transcripts in individuals infected with human immunodeficiency virus type 1 (HIV-1). Journal of General Virology.863–876.
Manghera M, Magnusson A, Douville RN. 2017. The sense behind retroviral anti-sense transcription. Virol J. 14(1):9.
Matsuoka M, Mesnard JM. 2020. HTLV-1 bZIP factor: The key viral gene for pathogenesis. Retrovirology. 17(1):2.
McManus WR, Bale MJ, Spindler J, Wiegand A, Musick A, Patro SC, Sobolewski MD, Musick VK, Anderson EM, Cyktor JC et al. 2019. HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. J Clin Invest. 129(11):4629-4642.
Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, Kingsley LA. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science. 272(5265):1167-1170.
Michael NL, Vahey MT, d'Arcy L, Ehrenberg PK, Mosca JD, Rappaport J, Redfield RR. 1994. Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by tat. J Virol. 68(2):979-987.
Miller RH. 1988. Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. Science. 239(4846):1420-1422.
Miura M, Yasunaga J, Tanabe J, Sugata K, Zhao T, Ma G, Miyazato P, Ohshima K, Kaneko A, Watanabe A et al. 2013. Characterization of simian t-cell leukemia virus type 1 in naturally infected japanese macaques as a model of HTLV-1 infection. Retrovirology. 10:118.
Pace M, Frater J. 2019. Curing HIV by 'kick and kill': From theory to practice? Expert Rev Anti Infect Ther. 17(6):383-386.
Pavesi A, Romerio F. 2022. Extending the coding potential of viral genomes with overlapping antisense ORFs: A case for the de novo creation of the gene encoding the antisense protein ASP of HIV-1. Viruses. 14(1).
Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. Nat Rev Genet. 14(12):880-893.
Perng GC, Jones C, Ciacci-Zanella J, Stone M, Henderson G, Yukht A, Slanina SM, Hofman FM, Ghiasi H, Nesburn AB et al. 2000. Virus-induced neuronal apoptosis blocked by the herpes simplex virus latency-associated transcript. Science. 287(5457):1500-1503.
Raffi F, Hanf M, Ferry T, Khatchatourian L, Joly V, Pugliese P, Katlama C, Robineau O, Chirouze C, Jacomet C et al. 2017. Impact of baseline plasma HIV-1 RNA and time to virological suppression on virological rebound according to first-line antiretroviral regimen. J Antimicrob Chemother. 72(12):3502.

Rasmussen MH, Ballarin-Gonzalez B, Liu J, Lassen LB, Fuchtbauer A, Fuchtbauer EM, Nielsen AL, Pedersen FS. 2010. Antisense transcription in gammaretroviruses as a mechanism of insertional activation of host genes. J Virol. 84(8):3780-3788.
Rosikiewicz W, Makalowska I. 2016. Biological functions of natural antisense transcripts. Acta Biochim Pol. 63(4):665-673.
Sanders-Buell E, Salminen M, McCutchan F. 1995. Sequencing primers for HIV-1. Human retroviruses and AIDS.
Satou Y, Yasunaga J, Yoshida M, Matsuoka M. 2006. Htlv-i basic leucine zipper factor gene mRNA supports proliferation of adult t cell leukemia cells. Proc Natl Acad Sci U S A. 103(3):720-725.
Schifano JM, Corcoran K, Kelkar H, Dittmer DP. 2017. Expression of the antisense-to-latency transcript long noncoding RNA in kaposi's sarcoma-associated herpesvirus. J Virol. 91(4).
Sengupta S, Siliciano RF. 2018. Targeting the latent reservoir for HIV-1. Immunity. 48(5):872-895.
Stevens JG, Wagner EK, Devi-Rao GB, Cook ML, Feldman LT. 1987. RNA complementary to a herpesvirus alpha gene mrna is prominent in latently infected neurons. Science. 235(4792):1056-1059.
Torresilla C, Mesnard JM, Barbeau B. 2015. Reviving an old HIV-1 gene: The HIV-1 antisense protein. Curr HIV Res. 13(2):117-124.
Global hiv & aids statistics — 2020 fact sheet. [accessed 12/21/2020]. https://www.unaids.org/en/resources/fact-sheet.
Van Zyl GU, Katusiime MG, Wiegand A, McManus WR, Bale MJ, Halvas EK, Luke B, Boltz VF, Spindler J, Laughton B et al. 2017. No evidence of hiv replication in children on antiretroviral therapy. J Clin Invest. 127(10):3827-3834.
Vanhee-Brossollet C, Thoreau H, Serpente N, D'Auriol L, Levy JP, Vaquero C. 1995. A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. Virology. 206(1):196-202.
Vansant G, Bruggemans A, Janssens J, Debyser Z. 2020. Block-and-lock strategies to cure hiv infection. Viruses. 12(1).
Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM. 2014. HIV latency. Proliferation of cells with hiv integrated into cancer genes contributes to persistent infection. Science. 345(6196):570-573.
Wang QY, Zhou C, Johnson KE, Colgrove RC, Coen DM, Knipe DM. 2005. Herpesviral latency-associated transcript gene promotes assembly of heterochromatin on viral lytic-gene promoters in latent infection. Proc Natl Acad Sci U S A. 102(44):16055-16059.
Wiegand A, Spindler J, Hong FF, Shao W, Cyktor JC, Cillo AR, Halvas EK, Coffin JM, Mellors JW, Kearney MF. 2017. Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during art. Proc Natl Acad Sci U S A. 114(18):E3659-E3668.
Wight M, Werner A. 2013. The functions of natural antisense transcripts. Essays Biochem. 54:91-101.

Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, Richman DD. 1997. Recovery of replication-competent hiv despite prolonged suppression of plasma viremia. Science. 278(5341):1291-1295.
Zapata JC, Campilongo F, Barclay RA, DeMarino C, Iglesias-Ussel MD, Kashanchi F, Romerio F. 2017. The human immunodeficiency virus 1 ASP RNA promotes viral latency by recruiting the polycomb repressor complex 2 and promoting nucleosome assembly. Virology. 506:34-44.
Zinad HS, Natasya I, Werner A. 2017. Natural antisense transcripts at the interface between host genome and mobile genetic elements. Front Microbiol. 8:2292.